Gender Gaps in the Evaluation of Research: Evidence from Submissions to Economics Conferences*

Laura Hospido
Bank of Spain and IZA

Carlos Sanz
Bank of Spain

February 3, 2020

Abstract

We study gender differences in the evaluation of submissions to economics conferences. Using data from the Annual Congress of the European Economic Association (2015-2017), the Annual Meeting of the Spanish Economic Association (2012-2017), and the Spring Meeting of Young Economists (2017), we find that all-female-authored papers are 3.3 p.p. (6.8%) less likely to be accepted than all-male-authored papers. This gap is present after controlling for number of authors of the paper; referee fixed effects; field; cites of the paper; authors' previous publication record, affiliations, and experience; and connections between the authors of a given paper and the referees that evaluate it. We provide evidence suggesting that the gap is driven by stereotypes against female authors: it is entirely driven by male referees, only exists for lesser-known authors, and seems larger in more masculine fields, especially in finance.

Keywords: gender, economics profession, academic labor market

JEL codes: A1, J16

*Laura Hospido: laura.hospido@bde.es. Carlos Sanz: carlossanz@bde.es. We thank Jaime Alonso-Carrera, Emmanuelle Auriol, Antonio Cabrales, Vera Z. Eichenauer, Irene Ferrari, Manu García, Nezih Guner, Eliana La Ferrara, Juan Francisco Jimeno, Gemma Prunner-Thomas, and John Rust for their help in obtaining the data; Chiara Giannetto, Víctor Gundín, Camila Mishell Pineda, and Rubén Suárez for outstanding research assistance; and Manuel Arellano, Olympia Bover, Paula Bustos, Nicola Fuchs-Schündeln, Thomas Fujiwara, Martín Gonzalez-Eiras, Nezih Guner, Andrea Ichino, Nagore Iriberri, Margarita Machelett, Monica Martínez-Bravo, Luigi Minale, Barbara Petrongolo, Roberto Ramos, Pau Roldán, Ernesto Villanueva, two anonymous referees, and seminar participants at the AEET Labour Economics Meeting, Annual Congress of the European Economic Association, Annual Meeting of the Spanish Economic Association, Bank of Spain, Conference on Gender and Career Progression, COSME Gender Economics Workshop, Universidad Complutense de Madrid, Universitat de Barcelona/IEB, Universitat de Girona, and Workshop on Public Policies: Inequality of Opportunity for useful comments. Special thanks to Manu García for invaluable programming help. The opinions and analyses are the responsibility of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

1 Introduction

Improving gender equality in academia and research is at the center of public policy debate, and is therefore attracting much attention in the academic literature. Economics in particular remains a male-dominated field. Although the share of women in economics has grown, it is still lower than in STEM fields.[1] For instance, in the US, women account for 33% of new PhDs, 29% of assistant professors at Ph.D. granting departments, and only 14% of full professors (Committee on the Status of Women in the Economics Profession (2017)). Potential explanations for the low representation of women include gender differences in the preferences for competitive environments (Niederle and Vesterlund (2007), Buser, Niederle, and Oosterbeek (2014)) or in bargaining abilities in the labor market (Blackaby, Booth, and Frank (2005)), child-rearing responsibilities (Bertrand (2013); Bertrand, Black, Jensen, and Lleras-Muney (2018)), and gender-based discrimination (Goldin and Rouse (2000)).

In this paper, we study gender differences in the evaluation of submissions to economics conferences. Conferences are an essential part of academic life. They are useful to receive feedback, improve presentation and communication skills, get to know fellow economists in the field, hear about the latest research, gain visibility, and develop networking and future collaborations. Indeed, it has been shown that conferences increase the cites of the articles presented there (de Leon and McQuillin (2018)) and the likelihood of co-authoring with another attendant in the future (Campos, Leon, and McQuillin (2018)). Hence, the presence of gender gaps in the evaluation process may have substantial impact on the professional careers of economists. Furthermore, understanding whether or not there are gender differences in this context can be informative about the state of gender equality in the academic environment.[2]

Conference submissions are evaluated through blind peer-review, which is an established component of professional practice. The fundamental principle is straightforward: experts in a given domain appraise the professional performance, creativity, or quality of scientific work produced by others in their field or area of competence (Lee, Sugimoto, Zhang, and Cronin (2013)). Peer review in economics covers a wide spectrum of activities, including the evaluation of research grant applications, review of articles submitted to journals, rating of papers and posters submitted to conferences, and promotion and tenure decisions. Threats to the impartiality of the review may be larger in the context of conferences, where referees evaluate a large number of papers in a short period of time. While there is some work on gender gaps at the evaluation of grants proposals, papers submitted to journals, and promotions, less is

[1] Lundberg and Stearns (2019), using data on the share of female faculty in top-50 departments for several science and social science disciplines from 2002 to 2012, document that economics remains within the lowest group, along with physics, math, and engineering, and far below the biological and other social sciences. Indeed, Ceci, Ginther, Kahn, and Williams (2014) find economics to be an "outlier" among academic fields because of "a persistent sex gap in promotion that cannot readily be explained by productivity differences".

[2] A recent survey (American Economic Association (2019)) describes a competitive climate in the economics profession that is hostile to women and, in particular, related to experiences of discrimination in invitations to participate in research conferences, associations, and networks.

known regarding conferences.

We have obtained unique data from the submissions to three general-interest academic conferences: the *Annual Congress of the European Economic Association* (EEA), the *Annual Meeting of the Spanish Economic Association* (SAEe), and the *Spring Meeting of Young Economists* (SMYE).[3] These conferences are some of the largest in the world. For example, in 2017 they hosted approximately 1,000, 350, and 150 presentations, respectively. Our dataset covers all submissions from 2015–2017 for the EEA, 2012–2017 for the SAEe, and 2017 for the SMYE, adding up to 9,342 submissions. It contains information on the gender of both the authors and the referees, and on the acceptance decision. In addition, we have complemented the data with a rich set of controls—the field, cites, and eventual publication of the paper, the rank of the affiliations, prior publication, and experience of the authors, and the connections between the authors and the referees that evaluate any given paper.

We begin by showing that a 1-p.p rise in the share of male authors is associated with a .054-p.p. rise in the probability that the paper is accepted, i.e. switching from an all-female-authored to an all-male-authored paper increases the probability of acceptance by 5.4 p.p. Given the baseline rate of acceptance for papers with all male authors (47.1%), this amounts to an 11.5% effect.

We then study whether this gap can be explained by several factors that may correlate with gender and acceptance rates. First, we control for the number of authors of the paper, given that women are more likely than men to single-author. Second, to account for the possible non-random assignment of papers to referees, we add referee fixed effects. Third, we include fifteen field fixed effects, as women are relatively more represented in some fields than others. Fourth, to account for the quality of the paper, we control for cites of the paper at submission year or, alternatively, ex-post cites and eventual publication of the paper. Fifth, to further control for quality, we include some characteristics of the authors: previous publication record, rank of their affiliations, and experience (years since they obtained their PhD). Finally, as connections may play a role in the evaluation process, we include two measures of the network connections between the authors of any given paper and the referees that evaluate it: (i) whether they are affiliated with the same institution (or, alternatively, with an institution based in the same city), and (ii) the "shortest path" in co-authorship between the referee and the authors (i.e. 1 if they are co-authors, 2 if they are co-authors of co-authors, etc.). Overall, we find that these factors are strongly correlated with the acceptance decision. For example, being affiliated with a top-50 institution increases the probability of acceptance by almost 30 p.p. Connections also have a large impact on acceptance: being in the same institution as the referee increases the probability of having the paper accepted by

[3]For this project, we contacted the European Economic Association, the Spanish Economic Association, the European Association of Young Economists, the American Economic Association, the Econometric Society, and the Royal Economic Society to obtain data from their general-interest conferences under confidentiality agreements. Concrete progress has been made for the first three.

14.1 p.p., and being a co-author (a co-author of a co-author) of the referee does so by 21.6 (14.1) p.p. However, after taking all of these factors into account, the gender gap is still sizable (3.3 p.p., or 6.8%) and statistically significant.

In addition, we show that the results are robust to a number of alternative exercises. The gap is similar in size across the three conferences. It is also not sensitive to alternative ways of specifying gender and the control variables. We also show suggestive evidence that omitted factors cannot fully account for the gender gap, by performing Oster (2019)'s sensitivity test to omitted variables.

We then discuss the possible mechanisms behind the remaining gap. Following arguments in Bagues, Sylos-Labini, and Zinovyeva (2017), we perform three sets of tests to study if the gap may be driven by stereotypes against female economists.[4] First, given that there is evidence that men may hold more negative stereotypes of women than women do, finding a larger gap when papers are evaluated by male referees will be suggestive of the presence of stereotypes.[5] We test this hypothesis and find that the gap is *entirely* driven by male referees: female referees evaluate male and female-authored papers similarly, but male referees are more favorable towards papers written by men. Second, stereotyping should be weaker when referees are more informed about the authors' quality. Consistent with this, we find that the gender gap disappears when evaluating authors at top institutions or with a good record of publications. And third, stereotyping should be stronger in fields that are less feminized. We present evidence suggesting that the gap may be larger in more masculine fields—especially, in finance.

While these three sets of results are consistent with the gap being driven by stereotypes, there are other possible mechanisms, which we discuss next. First, our quality controls (cites of the paper and prior publications and affiliations of the authors) might be gender biased. Gender bias in the quality controls could account for the gender gap if this bias is *in favor of women*, which we think is unlikely in light of other work (Card, DellaVigna, Funk, and Iriberri (2019), Sarsons, Gërxhani, Reuben, and Schram (2020)).[6] Second, it could be that papers' unobserved characteristics (for example, substantive relative to methodological contributions) differ by gender, and conferences' referees value the characteristics of male-authored papers more. While we cannot rule out this explanation, it is also hard to reconcile with our heterogeneity results—it is not clear why those unobserved characteristics would only be valued by male referees, or would not matter for well-known authors. Third, it could be that there is taste-based

[4] Bagues, Sylos-Labini, and Zinovyeva (2017) study the presence of stereotypes in the evaluations of applications to associate and full professorships in Italy and Spain.

[5] For example, 34.5% of men in Europe believe that men make better political leaders than women do, while only 22.9% of women believe this. Similarly, 32.9% (18.9%) of men (women) believe that men make better business executives than women do. In the US, these figures are 24.3% (14.8%) regarding the first question, and 15.5% (8.0%) for the second. Source: World Values Survey Wave 6: 2010-2014. European countries included: Cyprus, Estonia, Germany, Netherlands, Poland, Romania, Slovenia, Spain, Sweden, and Ukraine.

[6] Academics believe that females receive about 6% fewer citations than males, holding constant the quality of their papers, according to a survey conducted by Card, DellaVigna, Funk, and Iriberri (2019). Sarsons, Gërxhani, Reuben, and Schram (2020) find that, conditional on quality, women are less likely to receive tenure the more they coauthor, while this is not the case for men.

discrimination against female economists. However, the result that the gap is only against lesser-known authors favors an interpretation based on stereotypes instead of one based on taste-based discrimination (Bohren, Imas, and Rosenberg (2019)).

This paper contributes to four strands of the literature. First, by providing a systematic analysis of the evaluation of submissions to conferences, we contribute to the growing literature on gender differences in the evaluation of research. Our paper is most closely related to a concurrent working paper by Chari and Goldsmith-Pinkham (2018). They use data from the 2016 and 2017 editions of the NBER Summer Institute and find no difference in acceptance rates by gender. Submitters to the NBER Summer Institute are, on average, probably better known than submitters to the conferences that we study. Given that we find no gap for well-known authors, our results are compatible with theirs.[7] Other literature on the existence of gender gaps in the evaluation of research has found mixed results. Blank (1991), conducting a randomized experiment at the American Economic Review, finds no differences in the acceptance rates of female-authored papers in single-blind and double-blind submissions. Similarly, Abrevaya and Hamermesh (2012) find no evidence of gender differences at the evaluations at an anonymous journal. By contrast, Broder (1993), Wennerds and Wold (1997), and Van der Lee and Ellemers (2015) find that grant proposals submitted by women are rated lower than those submitted by men. Finally, two papers find gender gaps in the evaluation of submissions to journals. Card, DellaVigna, Funk, and Iriberri (2019), using data from four leading economics journals, find that papers by all-female authors with comparable referee recommendations end up accumulating more cites than all-male author papers. This suggests that referees set a higher bar for all-female-authored papers or, alternatively, that all-female-authored papers show characteristics that are valued differently by the refereeing process. Hengel (2018) finds that female-authored papers have abstracts that are better written, in terms of clarity, than male-authored papers, and interprets this as evidence of tougher editorial standards for women and/or biased referee assignment.

Second, by providing evidence on the interaction between the authors' and the referees' gender, our paper also contributes to the literature on "in-group" biases, i.e. the preferential treatment of individuals of one's group. In-group biases have been found, for example, among Jewish and Arab judges in Israel (Shayo and Zussman (2011)) and among NBA referees, who tend to favor players of their own race (Price and Wolfers (2010)). Regarding in-group *gender* biases, the evidence is mixed. Boring (2017) and Mengel, Sauermann, and Zölitz (2018) find that both male and female students evaluate male instructors higher, suggesting the presence of *absolute* biases against women. Similarly, Card, DellaVigna, Funk, and Iriberri (2019) and Krawczyk and Smyk (2016) provide evidence that both women and men are harsher against papers written by women. De Paola and Scoppa (2015) find that female candidates

[7] However, Chari and Goldsmith-Pinkham (2018)'s focus is more on how women's representation has evolved over fields and time and do not control for quality, so it is not straightforward to compare our estimates with theirs.

are less likely to be promoted when the committee is composed exclusively of males, while the gender gap disappears with mixed-sex committees, suggesting the presence of male in-group bias. Sarsons, Gërxhani, Reuben, and Schram (2020) find that male HR personnel are more likely to hire in favor of men, and women in favor of women, suggesting in-group biases for both males and females. Bagues, Sylos-Labini, and Zinovyeva (2017), however, find that a larger number of women in evaluation committees does not increase the number of female candidates who qualify. Finally, Bagues and Esteve-Volart (2010) find that female applicants to the Spanish judiciary have lower chances of being hired when they are randomly assigned to an evaluation committee including women. Our findings indicate the presence of male but not female "in-group" bias, as male (female) referees are more (equally) favorable to papers written by male (female) authors.

Third, we contribute to the literature on gender stereotypes. Bohren, Imas, and Rosenberg (2019) find, in a field experiment using online evaluations, that women are discriminated against when evaluators have not seen many prior contributions of the author. However, after more evaluations, the direction of discrimination reverses: women's work is favored over men's. Using a dynamic model, they argue that this pattern is consistent with biased beliefs or stereotypes against women. In line with this, we find that there is no gender gap for well-known authors, but we do not observe a reversal, i.e. we do not find that female well-known authors are favored relative to male well-known authors. Bordalo, Coffman, Gennaioli, and Shleifer (2019), using lab experiments, show that stereotypes vary by category, with individuals tending to overestimate the performance of other men (women) in categories that are judged to be men-typed (women-typed). Similarly, Sarsons, Gërxhani, Reuben, and Schram (2020) find that women are predicted to perform worse than their male counterparts for male-stereotyped quizzes, while in female-stereotyped quizzes, women and men are given equal credit. We provide evidence that is consistent with these findings by showing that the gap seems to be larger in more masculine fields.

Fourth, our paper contributes to the literature that studies the role of network connections in evaluation processes. In other contexts, the literature has found that connections between reviewers and candidates increase the probability of being appointed a professor in France (Combes, Linnemer, and Visser (2008)) and Italy (De Paola and Scoppa (2015), Durante, Labartino, and Perotti (2011), Perotti (2002)), promoted in Spain (Zinovyeva and Bagues (2015)), or being awarded a grant in Sweden (Sandström and Hällsten (2007)). There is less evidence, however, on the value of connections for conference evaluations. We contribute to this literature by showing that, in our context, connections have a large impact on papers' acceptance.

The paper proceeds as follows. Section 2 provides background on the conferences used in the analysis. Section 3 describes the data. Section 4 presents the main results. Section 5 lays out robustness checks. Section 6 discusses the possible mechanisms behind our findings, providing some additional results. Section 7 concludes.

2 Background

We have obtained data from paper submissions to three major conferences: the European Economic Association Annual Congress (EEA), the Annual Meeting of the Spanish Economic Association (SAEe), and the Spring Meeting of Young Economists (SMYE).

The European Economic Association is an international scientific body, with membership open to all persons involved or interested in economics. The Annual Congress, which takes place at the end of August or early September, is one of the main events among the Association's activities. The first Annual Congress was held in Vienna in 1986. Since then, different major cities in Europe have hosted it. In addition to contributed sessions, the Annual Congress features a Presidential Address, two plenary lectures, invited paper sessions, and panel debates. In recent years, the Annual Congress has been held jointly with the European Meeting of the Econometric Society. We have data for the 2015, 2016, and 2017 editions of the conference, which took place in Mannheim, Geneva, and Lisbon, respectively.

The main goal of the Spanish Economic Association is the promotion and dissemination of scientific knowledge in economics through the organization of conferences, publications, and any other activity that furthers that aim. The annual conference brings together experts in all areas of economics every year in December since 1976. From 1976 to 2001 the conference, originally called Simposio de Análisis Económico, was organized by the Universitat Autònoma de Barcelona. Since 2001, the Simposio became the annual conference of the Spanish Association of Economics and is hosted by a different institution each year. Our sample includes data for years 2012–2017, when it was organized by the Universidad de Vigo (2012), CEMFI and Universidad Internacional Menéndez Pelayo (2013), Universitat de les Illes Balears (2014), Universitat de Girona (2015), Universidad del País Vasco (2016), and Barcelona Graduate School of Economics (2017).

The European Association of Young Economists aims to facilitate the interaction between young researchers in economics. Its main activity is the organization of the yearly meeting: the Spring Meeting of Young Economists (SMYE). The SMYE started in 1996 as a small-scale event for German Ph.D. students in Essen. With driving forces Uwe Dulleck and Achim Wambach, it expanded rapidly from an initial 30 participants to over 100 participants during the third edition in Berlin. By now, the Spring Meeting of Young Economists has become an international event, which has been organized in 11 different countries, and receives over 700 applications per year. Only non-tenured researchers under the age of 35 can submit papers. Co-authors of submitted papers, however, can be of any status or age. For the SMYE, we have data from the 2017 edition, when it was organized in Mallorca.

These three conferences follow a similar evaluation procedure. A program chair or board is responsible for the selection of papers. The board assigns papers to referees, which are given a few weeks to evaluate the papers.[8] In our sample, a median of 8 papers per referee are assigned, and 38% of papers are assigned to more than one referee. This assignment is mostly done by field of research. Referees then grade papers. The process is not double-blind, as the name of the authors usually appears on the manuscript. On top of this, depending on the configuration of the conference software in any given edition, referees may also see there the list of co-authors and who the submitter is. A justification of the assessment is normally encouraged but not required. When this is provided, it is usually very short. In the EEA and the SAEe, referees are informed of the targeted acceptance rate. This is not the case in the SMYE. Although the program committee makes the final selection, referees are influential. In particular, the rules of the SMYE stipulate that the final selection is based on the average score, i.e. no discretionary decisions made by the EAYE Board. There is no such rule in the SAEe or in the EEA, but it is standard practice to follow referees' recommendations. In Section 6, we show that cases in which the committee overrules referees are rare.

3 Data

In this section, we describe the variables used in the analysis. For more details on the procedure employed to build the dataset, please see the Dataset Construction Appendix.

Outcomes. Our main outcome is a dummy variable that takes the value of one if a given paper was accepted into the conference. In Table 1, which shows the summary statistics, we can see that the mean acceptance rate was 52%. For the SMYE, and for the EEA in 2016 and 2017, we also observe the grades given by referees to papers. The mean grade, on a 0-10 scale, is 5.73. Finally, at the SMYE each referee can nominate each paper that he or she reviews to be considered for the best paper award. Hence, as an alternative outcome variable for this conference, we consider the share of referees that nominate the paper for the award.[9] The mean of this variable is .07, i.e., on average, 7% of the referees that evaluate a paper recommend it for the award.

Authors' gender. We define the authors' gender in three alternative ways: (i) the share of male authors of the paper (which we use as the baseline), (ii) three dummies indicating whether the paper has a majority of male authors, half-male authors, or a majority of female authors, and (iii) a dummy indicating whether the submitter is male.[10] The mean share of male authors is 69%. The majority of male, half-male, and the majority of female authors categories represent 65%, 12%, and 24% of the

[8] At the SMYE, referees are given approximately five or six weeks to submit their grades. At the SAEe and the EEA, they are given around two weeks.

[9] We only observe the number of referees that nominate each paper, not whether each individual referee nominates the paper. Hence, our measure is the number of nominations obtained divided by the number of referees that evaluated the paper.

[10] Card, DellaVigna, Funk, and Iriberri (2019) focus on the gender of the author with the most publications (the so-called senior author). We cannot follow the same approach as we only have 2.1% of the observations with a senior female author.

observations, respectively. Finally, the mean share of male submitters is 67%, close to the mean share of male authors.

Referees' gender. We consider the gender of referees in some of our empirical analyses. Table 1 shows that 76% of evaluations were done by a male referee.

Fields. We classify papers into fifteen fields following the EEA submissions categories.[11] Consistent with previous work that has documented that women are more represented in some fields than in others (e.g. Dolado, Felgueroso, and Almunia (2012)), we also find that this is the case in our sample. This can be seen in Table A1, which shows the summary statistics by the gender of the authors. For example, the share of observations on macroeconomics is .22 (.15) among papers written by a majority of male (female) authors, while this share is .11 (.20) on applied microeconomics.

Cites of the paper. We have collected the number of Google Scholar cites to account for the quality of the submitted paper. Our variable Cites is defined as the asinh of the number of cites of the paper at the submission year.[12] Our main definition includes existing cites until the submission year to ensure that this variable is not a "bad control", i.e. an outcome of having been accepted into the conference. However, one may think that, given that it takes some time for papers to be cited, an ex-post measure (i.e. some years after the conference) may capture the quality of the paper better. To take this possibility into account, we have also collected ex-post cites—the cites that the paper had in March 2019. We find that both measures of cites are highly correlated with the acceptance decision but, in fact, the ex-ante measure is a better predictor of acceptance. Finally, as an additional quality measure, we have collected data on whether the paper had been published by March 2019 in any of the 35 high-impact journals considered by Card, DellaVigna, Funk, and Iriberri (2019)—the list of journals can be consulted on Table A2. In Table A1, we can see that majority-male-authored papers are, on average, more cited than majority-female-authored ones.

Publication record of the authors. As an additional, indirect measure of quality, we consider the publication record of the authors in the years before the conference. Our main variable, No. Publications, is the average number of publications in the set of 35 high-impact journals in the five years prior to the submission year. For robustness, we also consider the number of publications of the most prolific co-author, the number of publications *in top-five journals* in the five years before submission, and the number of publications in the mentioned 35 journals in the *10* years before the submission. Table A1 reveals that the average number of prior publications is higher in papers with a majority of male authors than in those with a majority of female authors.

[11] The fields used by the EEA to categorize submissions are: applied microeconomics, behavioral, development, econometrics, history, theory, environmental, finance, industrial organization, international, labor, law and economics, macroeconomics, political economy, and public economics.

[12] The asinh transformation is similar to the log but can accommodate zero citations. It is defined as follows: $asinh(x) = ln(x + (1 + x^2)^{1/2})$.

Institutions of the authors. We consider the quality of the authors' affiliation at the submission year, measured through the IDEAS/RePEc ranking of institutions. For multiple-authored papers, we take the average rank of the authors' affiliations, showing robustness to considering instead the rank of the author in the highest-ranked institution. Our variables are six mutually-exclusive dummies indicating whether the affiliation is among the top-50 institutions, between the top 50 and the top 100, top 100-200, top 200-500, top 500-758 (which coincides with top the 10% institution), or below the top 10%.[13] For robustness, we also use an alternative ranking—the QS World University Ranking.[14] We do not observe large gender differences in the authors' affiliations.

Experience of the authors. To measure the experience of the authors, we have collected the year in which they obtained their PhD. For multiple-authored papers, we consider the average year (and show robustness to using the experience of the most experienced co-author). Our variables are six mutually-exclusive dummies indicating the number of years passed between the PhD year and the year of the conference—hence, negative values indicate that the co-author was a PhD student at the conference. As it can be seen in Table A1, women in our sample are, on average, less experienced than men. For example, observations with 15 or more years of experience are 12% in papers written by a majority of male authors, and 6% in papers written by a majority of female authors.

Connections between the authors and the referees. We have collected two measures of network connections. First, we have coded whether the referee and the authors of any given paper were affiliated with the same institution (or, alternatively, with an institution based in the same city) at the submission year. As with the rest of variables, for the baseline we consider the average, i.e. the share of co-authors in the same institution as the referee, and show robustness with a dummy indicating whether there was at least one co-author in the same institution as the referee. Second, we have accessed the webpage CollEc, which provides "shortest paths" in co-authorship between any pair of the over 43,000 economists in the RePEc network, to obtain shortest paths for referee-author pairs in our sample (as of October 2019). That is, if two economists are co-authors, then the shortest path equals one. If they are co-authors of co-authors, then the shortest path equals two, etc.[15] Table A1 shows that the shortest path is slightly lower for men than for women.

[13] RePEc does not specify the exact rank of institutions below the top 5%, just indicating whether they are between the top 5% and top 6%, between the top 6% and top 7%, etc. For these cases, we simply consider the mean rank of universities in that interval, e.g. if a university is between the top 5% (379 institutions) and the top 6% (454), we assign it rank 416.

[14] The advantage of this ranking is that, unlike the IDEAS/Repec ranking, it is not based on authors that register in the system. The drawback is that only universities are ranked.

[15] A limitation of this variable is that, given that some individuals are not registered in Repec, it is missing for 44% of our observations. The vast majority of these are weak connections, as individuals not registered in RePEc are, on average, young or not very active in research. In the analysis, we treat missing values as a separate category and, indeed, we observe that the probability of acceptance of authors in this category is similar to that for authors that are at a shortest path of more than 4. Furthermore, the presence of these missing values is unlikely to drive the results—the gender gap is also present (and is in fact larger) in the subsample with no missings.

4 Empirical Analysis

4.1 Unconditional Gap

Our research question is whether the authors' gender affects the probability that a paper is accepted into conferences. We begin by considering the following linear probability model:

$$\text{Accepted}_{prcy} = \beta \text{Share Male Authors}_{prcy} + \alpha_{cy} + \epsilon_{prcy}, \quad (1)$$

where Accepted is a dummy variable that takes the value of one if paper p, evaluated by referee r, submitted in year y to conference c, was accepted into the conference, Share Male Authors is the proportion of male authors in the paper, α_{cy} are conference-year fixed effects, β is the parameter of interest, and ϵ_{prcy} is an error term. The unit of observation is a pair paper-referee.

The results of fitting this model are in column (1) of Table 2. We see that a 1-p.p rise in the share of male authors in the paper is associated with a .054-p.p. rise in the probability that the paper is accepted, i.e. switching from an all-female-authored to an all-male-authored paper increases the probability of acceptance by 5.4 p.p. Given the baseline rate of acceptance for papers with all male authors (47.1%), this amounts to an 11.5% effect. This effect is significant at the 1% level.

4.2 Conditional Gap

In this subsection, we consider several potential explanations for the gap that we have documented. We sequentially add factors that may correlate with both the authors' gender and the acceptance decision, and see if and how much they can account for the observed gap.

Number of authors. It has been shown that women single-author more than men (see, e.g., Boschini and Sjögren (2007)). We find that this is also the case in our sample: the mean share of male authors for single-authored papers is .66, while it is above .71 for multiple-authored papers.[16] If referees are harsher evaluating single-authored papers, this may make female-authored papers less likely to be accepted. To account for this possibility, we add number-of-authors fixed effects to equation (1). The results, reported in column (2) of Table 2, reveal that controlling for the number of authors reduces but does not eliminate the gender gap, which is now 4.7 p.p.

Non-random assignment of papers to referees. It might be that female-authored papers are assigned to harsher referees, thus potentially creating the gender gap. To test this explanation, we add referee fixed effects to equation (1). The results, shown in column (3) of Table 2, indicate that the gender gap remains similar (5 p.p.) after doing so.

[16] More specifically, the mean share of male authors is .66 for papers with one author, .71 for papers with two and three authors, .74 for papers with four authors, .72 for papers with five authors, and .81 for papers with six authors.

Field. As discussed in the previous section, women are relatively more represented in some fields than others. If it is relatively harder to be accepted in more feminized fields (for example, because there are relatively fewer slots at conferences), then this might explain the gender gap. To take this issue into account, we add field fixed effects. Note, however, that the referee fixed effects most likely already account for this, as papers are assigned to referees mainly by topic. In fact, the results (column (4) of Table 2) reveal that the gender gap is also not sensitive to the inclusion of field fixed effects.

Cites of the paper. If women submit papers of lower quality than men, this might explain why the probability of acceptance of female-authored papers is lower. To control for quality, we add the cites of the paper as a control variable. The results, shown in column (5) of Table 2, indicate that this variable is highly correlated with the acceptance decision: as expected, more cited papers are more likely to be accepted. The gender gap after including this control is reduced to 4.5 p.p., but is still significant.

Publication record of the authors. As an additional measure of quality, we add as a control the average number of publications of the authors in a set of 35 leading journals in the five years before the submission. In column (6) of Table 2, we can see that one more prior publication increases the probability of acceptance by 7.8 p.p. After including this control, the gender gap is 3.1 p.p. (significant at the 5% level).

Institutions of the authors. As another measure of quality, we add the institution dummies as controls. We find that papers written by authors from a top 50 (50-100, 100-200, 200-500, 500-10%) institution are 32.0 p.p. (27.6, 24.2, 16.7, 9.3 p.p.) more likely to be accepted than those written by authors from a below-10% institution. These sizable effects may have two explanations. First, referees may use the affiliation of the authors as an element of judgment, i.e. they take it into consideration in the evaluation. Second, if the two previous quality controls (cites and previous publications) do not completely account for the quality of the paper, affiliations may capture some of it. After including these controls, the gender gap remains similar, at 3.2 p.p. (significant at the 5% level).

Experience of the authors. The fact that, as discussed above, female authors are, on average, less senior than male authors could create a gender gap if referees take experience into account when evaluating papers. In column (8) of Table 2, we add the experience dummies to the regression. Overall, we do not find large differences in the probability of acceptance by experience. If anything, it would seem that papers written by more experienced authors (more than 10 years) are less likely to be accepted. The gender gap remains unchanged when these controls are included.

Connections of authors and referees. Finally, if network connections play a role in the evaluation of papers, and male economists are better connected than female economists, this could generate the gender gap in acceptance rates.[17] By connections, we mean that the referee and the author(s) of the

[17] Regarding the first point, there is evidence that connections play an important role in evaluation processes, including national qualification exams, evaluations at the university level, and grant peer-review, as summarized in the introduction.

evaluated paper have a personal bond.[18] Column (9) of Table 2 provides the results when adding our two sets of connections variables to the regression. The results indicate that both are highly correlated with acceptance. A paper evaluated by a referee that is in the same institution as all the authors of the paper is 14.1 p.p. more likely to be accepted relative to one in which no author is in the same institution as the referee. Relative to missing paths, a shortest path of 1 (that is, all authors of the paper are coauthors the referee) increases the probability of acceptance by 21.6 p.p. Note that these large magnitudes are present conditional on a substantial number of factors, in particular, the quality of the institution. This suggests that connections are a strong driver of the evaluation process. However, the results also indicate that they cannot explain the gender gap. The estimated coefficient barely changes with the inclusion of connections to the regression, remaining at 3.3 p.p. (significant at the 5% level). This represents a 6.8% effect.

In sum, the results indicate that a non-negligible gender gap remains even after controlling for a considerable number of possible factors. In the next section, we assess the robustness of this result.

5 Robustness

In this section, we show that our results are robust to a number of robustness tests accounting for selection, potential omitted variables, and definition of variables. We begin by discussing the possibility of self-selection into submitting papers to conferences and what implications it may have on our estimates. We then address the issue of possible omitted variable bias, performing Oster (2019)'s test. Next, we test if the results are driven by a specific conference. Finally, we study the robustness to alternative definitions of the gender and control variables.

Selection. If there is self-selection into submitting papers to conferences, the characteristics of the authors in our sample could differ relative to the overall population of economists. For example, if lesser-known women believe that, conditional on submission, there is a gender gap against them (consistent with what we find in this paper), they may be less likely to submit their papers. This would make female applicants to conferences to be, on average, more well-known. On the other hand, several departments are trying to get gender balance in the invited seminar speakers. This might lead well-known female economists to be invited to present and less likely to apply to conferences given that they have already several opportunities to present. This would create self-selection in the other direction, that is, female economists that apply to conferences would be, on average, less well-known.

Regarding the second point, there is evidence that male economists are better connected than female economists. For example, Ductor, Goyal, and Prummer (2018) show that women have fewer collaborators than men. Hilmer and Hilmer (2007), observe that, in the US, around half of the economics PhD students being advised by women are female, while only 18% of economics PhD students advised by men are female.

[18] Cases in which this is likely to happen, and that are usually the focus of the literature on networks, are that the referee and the authors are or have been co-authors, colleagues, advisors, or mentors.

Overall, we observe that the pool of presenters in our sample is, in terms of gender, similar to the average in the profession. According to Auriol, Friebel, and Wilhelm (2019), the share of female positions in economics research institutions in Europe is 34.3% (32.8% in top-100 institutions).[19] In our sample, the share of female authors is 31%. It could still be the case that, although the gender composition in our sample is representative of the overall population of economists, there are gender differences by quality, i.e. that there is an over-representation of female authors of high quality and an under-representation of female authors of lower quality (or the opposite). Given that all our results are conditional on submission, this would only affect our results if it induces some quality differences that the included controls do not fully capture. We discuss this possibility next.

Omitted variables. Although we control for a considerable number of variables, the presence of some unobservable factors is obviously hard to rule out in this type of studies.[20] Here we discuss if omitted factors are likely to create the observed (conditional) gap in light of two pieces of evidence: (i) Oster (2019)'s sensitivity test and (ii) the heterogeneity results presented in Section 6.

(i) As a way to assess the robustness of our results to possible omitted variable bias, we perform Oster (2019)'s sensitivity test. This test is based on the idea proposed by Altonji, Elder, and Taber (2005) that if the estimated coefficient is robust to the inclusion of observable controls, then this is also an indication that it would not be affected much by the failure to control for other possible non-observable factors. Oster (2019)'s version takes into account the explanatory capacity of the covariates.[21]

In Table 3, we report the results. The bias-adjusted coefficients tell us what the effects would be under the assumption that unobservables have the same explanatory power than observables. For the specification with all the controls (columns (7)-(9)), the bias-adjusted coefficient is still positive and sizable (.020, vs. .033 in the baseline). We also report Oster's δ coefficients, which indicate how much more explanatory power should unobservables, relative to observables, have to fully account for the estimated effects. Oster (2019) suggests that δ values above one are reassuring about the robustness of the results. For the specification with all the controls, we obtain $\delta = 2.29$, indicating that possible omitted factors should have considerably more explanatory power than observables to account for the gap.

(ii) The heterogeneity results that will be presented in Section 6 also suggest that the gap cannot be fully explained by the existence of omitted factors. For example, unobserved quality cannot explain why the gap is entirely driven by male referees—unless one is willing to assume that male referees care more

[19]In the US, the share of female faculty in economics departments with doctoral programs is 23.5%, and 33.7% in those without them (Committee on the Status of Women in the Economics Profession (2017)).

[20]These factors could be related to self-selection, as discussed in the previous subsection, but could also exist in the absence of self-selection, if there are gender differences in the overall population of economists.

[21]Intuitively, it is more reassuring if the estimated coefficient remains invariant after controlling for covariates that raise a lot the R^2 than if this happens when controlling for covariates with less explanatory capacity.

about quality than female referees.

Results by conference. If the effect comes from a specific conference, one could worry that it is due to some idiosyncratic component specific to that conference, as opposed to revealing a general pattern in the profession. Hence, finding that the effect is similar in the three conferences would be reassuring about the external validity of the results.

We investigate whether the effect differs by conference by adding the interactions terms Sh. Male Authors x SAEe and Sh. Male Authors x SMYE to equation (1). The results are in column (1) of Table 4. We can see that on average, males are accepted at the EEA (the omitted category) 3.2 p.p. more than females. However, this estimate is rather imprecise due to the reduced number of observations when studying one subsample. The interactions terms are very close to zero and non-significant, suggesting that there is a gender gap of a similar magnitude in the three conferences.

Definition of the authors' gender. In column (2) of Table 4, we account for possible non-linearities in the effect of the share of male authors. We do this by replacing the share of male authors with dummies for having a majority of male authors and half-male authors, where papers with a majority of female authors are the omitted category. A gender gap is also observed under this alternative categorization of the authors' gender: papers with a majority of male authors are 3.1 p.p. more likely to be accepted than those with a majority of female authors, and this effect is significant at the 5% level. Papers co-authored by a team of half men and half women are equally likely to be accepted than papers with a majority of female authors.

In column (3) of Table 4, we separate the gender of the submitter and non-submitters.[22] We can see that there is a gender gap in both categories: papers submitted by a female are 2.6 p.p. less likely to be accepted than those submitted by men (conditional on the gender composition of non-submitters) and switching from a paper with no female non-submitters to one with all female non-submitters does so by 5.3 p.p. (conditional on the gender of the submitter).

Definition of the control variables. Finally, Table 5 shows how the estimates change if we control for: (i) ex-post (instead of ex-ante) number of cites, (ii) publication of the paper, (iii) number of publications in top-five (instead of top-35) journals, (iv) number of publications in the last ten (instead of five) years, (v) publications of the most prolific co-author (instead of average publications of the authors), (vi) publications based on the QS World University Ranking (instead of the IDEAS/RePEc ranking), (vii) experience of the most experienced co-author (instead of the average), (viii) closest shortest path (instead of the average), (xix) dummy for whether the referee and at least one of the authors are at the same institution (instead of fraction of authors at the same institution as the referee), (x) fraction of the authors

[22] As explained in Section 2, while we do observe who the submitter is, it is not clear if referees know this when evaluating the paper. The information is stored in the software but whether this information is observable by referees depends on the configuration of the software in the particular edition of the conference.

at an institution in the same city as the referee's, and (xi) dummy for whether the referee and at least one of the authors are in the same city. The gender gap remains remarkably stable across all specifications.[23]

6 Discussion of the Potential Mechanisms

In section 4, we saw that several potential factors (number of authors, non-random assignment of papers to referees, fields, cites of the paper, publication record, affiliation, and experience of the authors, and connections between the authors and the referees) explain some of the unconditional gap in acceptance rates. However, a gap of 3.3 p.p. remains even after taking these factors into account. In this section, we discuss four possible mechanisms behind the remaining gap.

Referees have stereotypes against female economists. Stereotypes are over-generalized beliefs about a particular category of people. As modeled by Bordalo, Coffman, Gennaioli, and Shleifer (2019), in the presence of stereotypes, individuals exaggerate group differences by focusing on the, often unlikely, features that distinguish one group from the other. In our context about gender, stereotypes would distort the perceived ability of the average member of a given gender. As argued by Bayer and Rouse (2016), stereotypes may be implicit or unconscious, which can produce behavior that diverges from the individual's own endorsed beliefs or principles.

Next, we perform three sets of tests, suggested in Bagues, Sylos-Labini, and Zinovyeva (2017), to study if stereotypes may play a role in explaining the observed gender gap.

(i) Given that there is evidence that men hold more negative stereotypes of women than women do, finding a larger gap in papers that are evaluated by male referees would suggest the presence of stereotypes against female economists.[24] To test whether this is the case, we add the interaction term Share Male Authors x Male Referee to equation (1). The results (column (1) of Table 6) indicate that the gender gap is entirely driven by *male* referees. When papers are evaluated by female referees, there is no gender difference between male and female-authored papers. When they are evaluated by male referees, there is a 4.5 p.p. gap (effect significant at the 10% level).

(ii) Stereotyping should be weaker when referees are more informed about the authors' quality. For example, in the extreme case that referees perfectly knew the authors' ability, they would be no role for stereotyping, as referees would not need to impute the authors' ability from the ability of the group. Hence, if the gap is driven by gender stereotypes, we expect that there is a smaller gap against well-known authors. We test for this by considering five alternative definitions of being "well-known": being

[23]The gender gap is also virtually unchanged if we consider a probit instead of a linear probability model (point estimate: .033; standard error: .010), or if we weight the observations by the inverse of the number of referees per paper (point estimate: .033; standard error: .014).

[24]There are gender differences on perceptions on whether men can be be better political leaders or business executives than women, as detailed in footnote 5. There is also some evidence that such differences could exist in academia. For example, editors of prominent economics journals surveyed by Card, DellaVigna, Funk, and Iriberri (2019), think that male reviewers are less likely to give a positive evaluation of female-authored papers than female reviewers.

affiliated with a top-100 or top-50 institution, or having published at least one paper in a top-35 journal in the last five or ten years before submission, or in a top-five journal in the last five years before submission.[25] We replace our treatment, Share Male Authors, with two variables that measure the share of male authors among well-known authors, Share Male Authors | Well-Known, and among lesser-known authors, Share Male Authors | Lesser-Known.[26] In these specifications, we include number-of-well-known-authors dummies on top of number-of-authors dummies. The results (columns (2)-(6) of Table 6) indicate that there is no gender gap against well-known authors. In the five specifications, the coefficient on Share Male Authors | Well-Known is close to zero and non-significant. All the gap is driven by lesser-known authors, with coefficients around 3-4.5 p.p. and statistically significant at the 1 or 5% levels, depending on specification.

(iii) Stereotyping should be stronger in fields that are less feminized. According to Bagues, Sylos-Labini, and Zinovyeva (2017), we should expect more stereotyping in more masculine fields because these fields offer fewer chances to interact with women.[27] Hence, if the gap is driven by gender stereotypes, we expect the gap to be larger in relatively more masculine fields. To test whether this is the case, we interact our treatment, Share Male Authors, with the variable Masculine Field, which takes the value of one (zero) if the share of male authors in the field of the paper is above (below) the median—or, alternatively, the 75th percentile.[28] The results, displayed in columns (7) and (8) of Table 6, suggest that the gap is higher in more masculine fields. The interaction is quantitatively large (but not statistically significant): the gender gap is 3.1 p.p. (three times) larger in fields with more men than the median, and 4.6 p.p. (3.7 times) larger in fields with more men than the 75th percentile. Finally, in column (9) of Table 6, we show the effects by macro-areas: empirical micro (the omitted category), macro, theory, and finance.[29] The results indicate that the gap is 1.6 p.p. in empirical micro, and 1.8 p.p. higher in macro

[25]These variables are, as expected, positively correlated, but not redundant. For example, among the set of authors with at least one paper in a top-35 journal in the last five or ten years before submission, 70% are not in top-100 institutions. Correlations between measures based on prior publications and measures based on affiliations are around .10. Within measures based on publications, the ones based on top-five and top-35 journals in the last five years also seem to capture different profiles: their correlation is .48. The time period, however, matters less: the correlation between top-35 publications in five versus ten years is .92.

[26]That is, for papers in which all authors are well-known, Share Male Authors | Well-Known = Share Male Authors. Similarly, for papers in which all authors are lesser-known, Share Male Authors | Lesser-Known = Share Male Authors. For papers with a mix of well-known and lesser-known authors, Share Male Authors | Well-Known = Number Male Well-Known Authors/Number Well-Known Authors, and Share Male Authors | Lesser-Known is defined analogously.

[27]Similarly, Bordalo, Coffman, Gennaioli, and Shleifer (2019) provide evidence that stereotypes vary by category, with individuals tending to overestimate the performance of other men (women) in categories that are judged to be men-typed (women-typed). To the extent that more masculine fields are more men-typed, i.e. men have some (probably small) advantage in performance in those fields, stereotypes would lead to a larger gap in more masculine fields.

[28]Fields above the median share of male authors are econometrics, theory, finance, industrial organization, law and economics, macroeconomics, political economy, and public economics. Fields above the 75th percentile are theory, finance, macroeconomics, and political economy.

[29]We assign fields to macro-areas as follows. Fields in the applied-micro area are applied micro, behavioral, development, geography, history, environmental, industrial organization, labor, law and economics, and public. Fields in the macroeconomics area are macroeconomics, international, and econometrics. Fields in the theory area are theory and political economy. The finance area coincides with the finance field. Assigning econometrics or behavioral to theory does not change the results.

and .9 p.p. higher in theory (but these differences are not statistically significant). The largest difference is with finance: the gap is 11.7 p.p. larger than in empirical micro (significant at the 5% level).[30]

In sum, there is evidence that the gap is driven by male referees, only exists for lesser-known authors, and seems larger in more masculine fields. All of these results are consistent with *referee* stereotypes against women. However, as mentioned in Section 2, the acceptance decision is potentially influenced not only by referees but also by the program committee. Although the fact that the gender gap remains unchanged when referee fixed effects are included suggests that referees drive the gap, here we provide additional evidence that reinforces this conclusion. We exploit that, for the SMYE and for the EEA in 2016 and 2017, we can observe the grade given by each referee to each paper. For the SMYE, we also have data on whether the paper was nominated by referees for the best paper award. Using this information, we first study if there is also a gender gap when we consider Grade or Nomination as the outcome. Column (1) of Table 7 shows, for reference, the effect on Accepted using the subsample in which Grade is available. Column (2) shows that all-male-authored papers receive, on a 0-10 scale, .14 more points than all-female-authored papers (effect significant at the 10% level), and column (3) shows that they are 2.6 p.p. more likely to be nominated (significant at the 5% level). Hence, there is also a gender gap when we look at referee-specific outcomes. Second, we check how frequent overrules are, i.e. cases in which a paper is accepted even though it should have been rejected based on the grade received by referees, or vice versa.[31] We find them to be rare—only 3.1% of cases—and uncorrelated with the authors' gender. In columns (4)-(6) of Table 7, we show that the results presented in columns (1)-(3) do not change when dropping overrules from the sample.[32]

Our quality measures are gender biased. One possibility is that the number of cites is gender biased. If this is the case, we may find a gender gap in the evaluation of submissions to conferences after controlling for quality, but this would not imply a gender *bias* in conferences' evaluations. Note that, under this explanation, for the evaluations of submissions to conferences to be gender neutral, it should be that the number of cites is biased *in favor of women*, i.e. female-authored papers are more cited than male-authored papers conditional on quality. If cites are biased *against* women, then the gender bias in conferences evaluations would be larger than what we find in this paper, i.e. we would be obtaining a lower bound of the bias.

[30]This is consistent with the situation in the finance profession described by Sherman and Tookes (2019). They study finance faculty from the top-100 US business schools during 2009-2017 and find that women tend to have positions at lower-ranked institutions, are less likely to be tenured, and are paid approximately 4% less than men.

[31]Suppose that, in a given conference-year, N papers are accepted. We define the cutoff grade as the grade obtained by paper ranked in position N. We say that there is an overrule if a paper with a grade lower (higher) than the cutoff grade is accepted (rejected).

[32]We have also estimated whether the gap differs by the gender of the chair of the admissions committee. The chair was female in the EEA in 2017, in the SAEe in 2012 and 2016, and in the SMYE. We cannot reject the null that the gap is the same with male and female chairs: the interaction term on Share Male Authors x Female Chair is close to zero and non-significant (point estimate: -.001; standard error: .019).

While determining biases in citations is beyond the scope of this paper, Card, DellaVigna, Funk, and Iriberri (2019) argue, based on a survey to economists, that citations are biased against women. If this is the case, we would be underestimating the gender bias in the evaluation of submissions to conferences.

Finally, note that the same argument applies to our other quality measures, namely, the quality of the authors' prior publications and affiliations. Gender biases in these variables could account for the gender gap, to the extent that they are biased in favor of women, i.e. women publishing better and being affiliated with better institutions than men of the same quality.[33]

Papers' unobserved characteristics differ by gender. It may be that female-authored papers lack some characteristics that are especially valued by conferences' referees. For example, if, in comparison with male-authored papers, female-authored papers make more substantive relative to methodological contributions, and conferences' referees are especially interested in this type of papers, this could generate a gender gap in acceptance rates.[34]

Although we cannot rule out this channel, we do not believe that it is the most natural explanation to our findings, as it cannot easily account for the observed heterogeneities. First, it is not clear why we should expect that these unobserved papers' characteristics are only valued by male and not female referees. Second, it is not clear how this mechanism can explain why there is only a gender gap for lesser-known authors.

Taste-based discrimination against female economists. Taste-based discrimination is consistent with the finding that the gap is driven by male referees. However, that the gap is only against lesser-known authors favors an interpretation based on stereotypes relative to one based on taste-based discrimination (Bohren, Imas, and Rosenberg (2019)).

7 Conclusion

Using data from three large European conferences, we find that all-female-authored papers are 3.3 p.p. (6.8%) less likely to be accepted than all-male-authored papers. This gap is present after accounting for a considerable number of factors, including cites of the paper, authors' publication record, affiliation, and experience, and connections between authors and referees. Our results indicate that the gender gap is entirely driven by male referees, is against lesser-known authors, and seems larger in more male-dominated fields, especially in finance, suggesting that the gap is driven by stereotypes against female economists.

These findings have direct policy implications for the design of systems to evaluate research and,

[33] However, Sarsons, Gërxhani, Reuben, and Schram (2020) suggests that women are given *less* credit for papers written with men, affecting negatively their tenure prospects.

[34] We say "especially" interested because, given that we control for fields and quality, these unobserved characteristics must be things that are valued by referees beyond field and quality (as proxied by the number of cites, prior publications, and institutions).

more specifically, to select papers for conferences. In particular, they imply that a more gender-balanced pool of referees would lead to more gender-neutral acceptance decisions. For example, the Executive Committee of the Spanish Economic Association has recently decided that future editions of the SAEe will have a gender-balanced panel of referees. Our results suggest that this decision may enhance equality of opportunities for female economists.

References

Abrevaya, J. and D. 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, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political economy 113*(1), 151–184.

American Economic Association (2019). AEA professional climate survey: Main findings. *The American Economic Association (AEA), available at: https://www.aeaweb.org/resources/member-docs/climate-survey-results-mar-18-2019*.

Auriol, E., G. Friebel, and S. Wilhelm (2019). Women in European economics. *VoxEU CEPR Policy Portal, available at https://voxeu.org/article/women-european-economics*.

Bagues, M. and P. Campa (2017). Can gender quotas in candidate lists empower women? Evidence from a regression discontinuity design. *CEPR Discussion Paper 12149*.

Bagues, M. and B. Esteve-Volart (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *The Review of Economic Studies 77*(4), 1301–1328.

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017). Does the gender composition of scientific committees matter? *American Economic Review 107*(4), 1207–38.

Bayer, A. and C. E. Rouse (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives 30*(4), 221–42.

Bertrand, M. (2013). Career, family, and the well-being of college-educated women. *American Economic Review 103*(3), 244–50.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2018). Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway. *The Review of Economic Studies 86*(1), 191–239.

Blackaby, D., A. L. Booth, and J. Frank (2005). Outside offers and the gender pay gap: Empirical evidence from the UK academic labour market. *The Economic Journal 115*(501), F81–F107.

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

Bohren, J. A., A. Imas, and M. Rosenberg (2019). The dynamics of discrimination: Theory and evidence. *American economic review 109*(10), 3395–3436.

Bordalo, P., K. Coffman, N. Gennaioli, and A. Shleifer (2019). Beliefs about gender. *American Economic Review 109*(3), 739–73.

Boring, A. (2017). Gender biases in student evaluations of teaching. *Journal of Public Economics 145*, 27–41.

Boschini, A. and A. Sjögren (2007). Is team formation gender neutral? Evidence from coauthorship patterns. *Journal of Labor Economics 25*(2), 325–365.

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

Buser, T., M. Niederle, and H. Oosterbeek (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics 129*(3), 1409–1447.

Campos, R., F. Leon, and B. McQuillin (2018). Lost in the storm: the academic collaborations that went missing in hurricane issac. *The Economic Journal 128*(610), 995–1018.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2019). Are referees and editors in economics gender neutral? *The Quarterly Journal of Economics, forthcoming*.

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), 75–141.

Chari, A. and P. Goldsmith-Pinkham (2018). Gender representation in economics across topics and time: Evidence from the NBER Summer Institute. *NBER Working Paper No. 23953*.

Combes, P.-P., L. Linnemer, and M. Visser (2008). Publish or peer-rich? The role of skills and networks in hiring economics professors. *Labour Economics 15*(3), 423–441.

Committee on the Status of Women in the Economics Profession (2017). Annual report. *The American Economic Association (AEA), available at: https://www.aeaweb.org/content/file?id=6388*.

de Leon, F. L. and B. McQuillin (2018). The role of conferences on the pathway to academic impact: Evidence from a natural experiment. *Journal of Human Resources*, 1116–8387R.

De Paola, M. and V. Scoppa (2015). Gender discrimination and evaluators gender: Evidence from Italian academia. *Economica 82*(325), 162–188.

Dolado, J. J., F. Felgueroso, and M. Almunia (2012). Are men and women-economists evenly distributed across research fields? Some new empirical evidence. *SERIEs 3*(3), 367–393.

Ductor, L., S. Goyal, and A. Prummer (2018). Gender and collaboration. *Working Paper, School of Economics and Finance, Queen Mary University of London*.

Durante, R., G. Labartino, and R. Perotti (2011). Academic dynasties: decentralization and familism in the Italian academia. *NBER WP 17572*.

Goldin, C. and C. Rouse (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review 90*(4), 715–741.

Hengel, E. (2018). Publishing while female. Are women held to higher standards? Evidence from peer review. *Unpublished Manuscript*.

Hilmer, C. and M. Hilmer (2007). Women helping women, men helping women? Same-gender mentoring, initial job placements, and early career publishing success for economics PhDs. *American Economic Review 97*(2), 422–426.

Krawczyk, M. and M. Smyk (2016). Author's gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. *European Economic Review 90*, 326–335.

Lee, C. J., C. R. Sugimoto, G. Zhang, and B. Cronin (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology 64*(1), 2–17.

Lundberg, S. and J. Stearns (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives 33*(1), 3–22.

Mengel, F., J. Sauermann, and U. Zölitz (2018). Gender bias in teaching evaluations. *Journal of the European Economic Association 17*(2), 535–566.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Perotti, R. (2002). The Italian university system: rules vs. incentives. In *first conference on Monitoring Italy, ISAE, Rome*.

Price, J. and J. Wolfers (2010). Racial discrimination among NBA referees. *The Quarterly Journal of Economics 125*(4), 1859–1887.

Sandström, U. and M. Hällsten (2007). Persistent nepotism in peer-review. *Scientometrics 74*(2), 175–189.

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2020). Recognition for group work: Gender differences in academia. *Unpublished manuscript*.

Shayo, M. and A. Zussman (2011). Judicial ingroup bias in the shadow of terrorism. *The Quarterly Journal of Economics 126*(3), 1447–1484.

Sherman, M. G. and H. Tookes (2019). Female representation in the academic finance profession. *SSRN Working Paper 3438653*.

Tang, C., K. Ross, N. Saxena, and R. Chen (2011). What's in a name: A study of names, gender inference, and gender behavior in Facebook. In *International Conference on Database Systems for Advanced Applications*, pp. 344–356. Springer.

Van der Lee, R. and N. Ellemers (2015). Gender contributes to personal research funding success in The Netherlands. *Proceedings of the National Academy of Sciences 112*(40), 12349–12353.

Wennerds, C. and A. Wold (1997). Nepotism and sexism in peer-review. *Nature 387*.

Zinovyeva, N. and M. Bagues (2015). The role of connections in academic promotions. *American Economic Journal: Applied Economics 7*(2), 264–92.

Tables

Table 1: Summary Statistics

	mean	min	max	sd	count
Accepted	0.52	0.00	1.00	0.50	16126
Grade	5.73	0.00	10.00	2.16	5825
Nomination	0.07	0.00	1.00	0.15	1949
Sh. Male Authors	0.69	0.00	1.00	0.40	16126
Half Male Authors	0.12	0.00	1.00	0.32	16126
Majority Male Authors	0.65	0.00	1.00	0.48	16126
Majority Female Authors	0.24	0.00	1.00	0.43	16126
Male Submitter	0.67	0.00	1.00	0.47	16111
Male Referee	0.76	0.00	1.00	0.42	16126
Number of Authors	1.77	1.00	6.00	0.90	16126
Applied Micro	0.14	0.00	1.00	0.35	16126
Behavioral	0.06	0.00	1.00	0.23	16126
Development	0.05	0.00	1.00	0.22	16126
Econometrics	0.04	0.00	1.00	0.19	16126
History	0.01	0.00	1.00	0.10	16126
Theory	0.06	0.00	1.00	0.24	16126
Environmental	0.01	0.00	1.00	0.11	16126
Finance	0.10	0.00	1.00	0.30	16126
IO	0.05	0.00	1.00	0.22	16126
International	0.07	0.00	1.00	0.26	16126
Labor	0.06	0.00	1.00	0.24	16126
Law and Economics	0.00	0.00	1.00	0.05	16126
Macroeconomics	0.20	0.00	1.00	0.40	16126
Political Economy	0.03	0.00	1.00	0.17	16126
Public	0.10	0.00	1.00	0.30	16126
Cites	0.13	0.00	5.43	0.46	16126
Cites Ex Post	0.99	0.00	6.68	1.35	16126
Published	0.08	0.00	1.00	0.28	16126
No. Publications	0.36	0.00	16.00	0.91	16126
No. Publications (10 Y)	0.62	0.00	20.00	1.56	16126
No. Publications (Top 5)	0.06	0.00	4.33	0.25	16126
Top 50 Ins.	0.11	0.00	1.00	0.31	16126
Top 50-100 Ins.	0.11	0.00	1.00	0.32	16126
Top 100-200 Ins.	0.19	0.00	1.00	0.39	16126
Top 200-500 Ins.	0.34	0.00	1.00	0.47	16126
Top 500-10% Ins.	0.09	0.00	1.00	0.29	16126
Below 10% Ins.	0.17	0.00	1.00	0.37	16126
Exp. ≤ 0	0.27	0.00	1.00	0.44	16126
Exp. $\in$ (0,5]	0.26	0.00	1.00	0.44	16126
Exp. $\in$ (5,10]	0.20	0.00	1.00	0.40	16126
Exp. $\in$ (10,15]	0.12	0.00	1.00	0.32	16126
Exp. $\in$ (15,20]	0.06	0.00	1.00	0.24	16126
Exp. > 20	0.04	0.00	1.00	0.21	16126
Exp. Missing	0.05	0.00	1.00	0.22	16126
Shortest Path ≤ 1	0.01	0.00	1.00	0.08	16126
Shortest Path $\in$ (1,2]	0.02	0.00	1.00	0.13	16126
Shortest Path $\in$ (2,3]	0.09	0.00	1.00	0.29	16126
Shortest Path $\in$ (3,4]	0.23	0.00	1.00	0.42	16126
Shortest Path > 4	0.22	0.00	1.00	0.41	16126
Shortest Path Missing	0.44	0.00	1.00	0.50	16126
Same Ins.	0.01	0.00	1.00	0.10	16126
Same City	0.02	0.00	1.00	0.13	16126

The unit of observation is a pair paper-referee. The variable Grade is available only for the SMYE and for the EEA in 2016 and 2017, and the variable Nomination is available only for the SMYE. In 15 observations the submitter information is missing.

Table 2: The Impact of the Authors' Gender on the Probability of Acceptance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sh. Male Authors	0.054***	0.047***	0.050***	0.048***	0.045***	0.031**	0.033**	0.033**	0.033**
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)
Cites					0.092***	0.078***	0.062***	0.063***	0.060***
					(0.013)	(0.013)	(0.012)	(0.012)	(0.012)
No. Publications						0.078***	0.065***	0.068***	0.060***
						(0.007)	(0.007)	(0.007)	(0.007)
Top 50 Ins.							0.320***	0.307***	0.295***
							(0.022)	(0.022)	(0.022)
Top 50-100 Ins.							0.276***	0.265***	0.251***
							(0.022)	(0.022)	(0.022)
Top 100-200 Ins.							0.242***	0.232***	0.224***
							(0.020)	(0.020)	(0.020)
Top 200-500 Ins.							0.167***	0.161***	0.157***
							(0.017)	(0.017)	(0.017)
Top 500-10% Ins.							0.093***	0.093***	0.090***
							(0.022)	(0.022)	(0.022)
Experience ≤ 0								0.124***	0.125***
								(0.025)	(0.025)
Experience $\in$ (0,5]								0.112***	0.108***
								(0.026)	(0.026)
Experience $\in$ (5,10]								0.118***	0.111***
								(0.027)	(0.028)
Experience $\in$ (10,15]								0.091***	0.082***
								(0.029)	(0.029)
Experience $\in$ (15,20]								0.090***	0.078**
								(0.031)	(0.031)
Experience > 20								0.077**	0.065*
								(0.034)	(0.034)
Shortest Path ≤ 1									0.216***
									(0.035)
Shortest Path $\in$ (1,2]									0.141***
									(0.030)
Shortest Path $\in$ (2,3]									0.076***
									(0.021)
Shortest Path $\in$ (3,4]									0.037**
									(0.017)
Shortest Path > 4									-0.024
									(0.016)
Same Institution									0.141***
									(0.037)
Constant	0.484***	0.489***	0.487***	0.477***	0.470***	0.454***	0.283***	0.183***	0.177***
	(0.013)	(0.013)	(0.011)	(0.023)	(0.023)	(0.023)	(0.026)	(0.033)	(0.034)
Observations	16126	16126	16126	16126	16126	16126	16126	16126	16126
R^2	0.10	0.11	0.17	0.17	0.18	0.19	0.23	0.23	0.24
Conf.-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
No. Authors FE		Y	Y	Y	Y	Y	Y	Y	Y
Referee FE			Y	Y	Y	Y	Y	Y	Y
Field FE				Y	Y	Y	Y	Y	Y

Results of regressing a dummy indicating whether the paper was accepted on the indicated set of variables and fixed effects. Cites is the asinh of the number of cites of the paper at the submission year, No. Publications is the average number of publications of the authors in the set of 35 high-impact journals specified in Table A2 in the five years prior to the submission year, the Top Ins. variables are dummies indicating the average rank of the affiliation of the authors (omitted category: below 10% institutions), the Experience variables are dummies indicating the average years passed since the authors obtained their PhD (omitted category: missing experience), the Shortest Path variables are dummies indicating the average shortest path in co-authorship between the referee and the authors (omitted category: missing shortest path), and Same Institution is the fraction of the authors that are in the same institution as the referee. Standard errors clustered by paper and referee in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Sensitivity Test: Coefficient Stability to Omitted Variables

		$a)$			$b)$			$c)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$\tilde{\beta}$	Adj. β	δ	$\tilde{\beta}$	Adj. β	δ	$\tilde{\beta}$	Adj. β	δ
Sh. Male Authors	0.048	0.042	5.26	0.033	0.019	2.24	0.033	0.020	2.29
No. Authors FE		Yes			Yes			Yes	
Referee FE		Yes			Yes			Yes	
Field FE		Yes			Yes			Yes	
Cites		No			Yes			Yes	
No. Publications		No			Yes			Yes	
Institutions		No			Yes			Yes	
Experience		No			No			Yes	
Connections		No			No			Yes	

Columns (1), (4), and (7) show the estimated effect, $\tilde{\beta}$ (also reported in Table 2), including different sets of control variables. Columns (2), (5), and (8) show the bias-adjusted β from Oster (2019) under the assumption that $\delta = 1$ and R_{max}=1.3$\tilde{R}$, with an upper limit of R_{max}=1. Columns (3), (6), and (9) show the value of δ that makes the coefficient zero.

Table 4: The Impact of the Authors' Gender on the Probability of Acceptance: Additional Results

	(1)	(2)	(3)
Sh. Male Authors	0.032		
	(0.020)		
Sh. Male Authors x SAEe	0.001		
	(0.028)		
Sh. Male Authors x SMYE	0.007		
	(0.042)		
Half Male Authors		-0.002	
		(0.024)	
Majority Male Authors		0.031^{**}	
		(0.014)	
Sh. Male Authors | Submitter			0.026^{**}
			(0.012)
Sh. Male Authors | Non-Submitters			0.053^{***}
			(0.020)
Constant	0.177^{***}	0.180^{***}	0.164^{***}
	(0.034)	(0.034)	(0.034)
Observations	16126	16126	16111
R^2	0.24	0.24	0.24

Results of regressing a dummy indicating whether the paper was accepted on the indicated set of variables. In addition to the reported coefficients, all columns contain the full set of controls and fixed effects included in column (9) of Table 2. Standard errors clustered by paper and referee in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: The Impact of the Authors' Gender on the Probability of Acceptance: Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Sh. Male Authors	0.031**	0.034**	0.039***	0.034**	0.031**	0.029**	0.033**	0.032**	0.033**	0.033**	0.033**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Cites Ex Post	0.046***										
	(0.005)										
Published	0.059***										
	(0.020)										
No. Publications (10 Y)		0.037***									
		(0.004)									
No. Publications (Top 5)			0.135***								
			(0.021)								
No. Publications (Max.)				0.033***							
				(0.003)							
Top 50 Ins. (Max.)					0.303***						
					(0.020)						
Top 50-100 Ins. (Max.)					0.248***						
					(0.021)						
Top 100-200 Ins. (Max.)					0.229***						
					(0.020)						
Top 200-500 Ins. (Max.)					0.157***						
					(0.019)						
Top 500-10% Ins. (Max.)					0.092***						
					(0.018)						
Top 50 Ins. (QS)						0.201***					
						(0.018)					
Top 50-100 Ins. (QS)						0.132***					
						(0.018)					
Top 100-200 Ins. (QS)						0.067***					
						(0.016)					
Top 200-400 Ins. (QS)						0.001					
						(0.019)					
Exp. ≤ 0 (Max.)							0.075***				
							(0.020)				
Exp. $\in$ (0,5] (Max.)							0.077***				
							(0.019)				
Exp. $\in$ (5,10] (Max.)							0.022				
							(0.019)				
Exp. $\in$ (10,15] (Max.)							0.020				
							(0.020)				
Exp. $\in$ (15,20] (Max.)							0.020				
							(0.022)				
Shortest Path=1 (Min.)								0.179***			
								(0.032)			
Shortest Path=2 (Min.)								0.162***			
								(0.026)			
Shortest Path=3 (Min.)								0.079***			
								(0.020)			
Shortest Path=4 (Min.)								0.023			
								(0.017)			
Shortest Path=5 (Min.)								-0.027			
								(0.017)			
Same Ins. (Max.)									0.101***		
									(0.028)		
Same City										0.121***	
										(0.030)	
Same City (Max.)											0.085***
											(0.024)
Observations	16126	16126	16126	16126	16126	16126	16126	16126	16126	16126	16126
R^2	0.25	0.24	0.23	0.24	0.24	0.23	0.24	0.24	0.24	0.24	0.24

Results of regressing a dummy indicating whether the paper was accepted on the full set of variables and fixed effects included in column (9) of Table 2, with the following changes: column (1) replaces Cites with Cites Ex Post and Published (a dummy indicating whether the paper has been published by March 2019 at any journal in the set of 35 high-impact journals specified in Table A2); columns (2) and (3) replace No. Publications with No. Publications (10 Y) and No. Publications (Top 5), respectively, which are publications in the ten (instead of five) years before the submission year, and publications in top-five (instead of top-35) journals; column (4) replaces No. Publications with No. Publications (Max.), which are publications by the most-prolific co-author (instead of average publications by all co-authors); column (5) replaces the institutions dummies based on the average quality of the authors' affiliations with institutions dummies based on the affiliation of the best-placed co-author; column(6) replaces the institution dummies (based on the RePEc/IDEAS ranking) with institution dummies based on the QS World University Ranking; column (7) replaces the experience dummies based on the average experience of the authors with experience dummies based on the experience of the most-experienced co-author; column (8) replaces the shortest-path dummies based on the average shortest path between the authors and the referee with shortest-path dummies based on the shortest-path of the co-author that is closest to the referee; columns (9)-(11) replace Same Institution with Same Ins. (Max.)—a dummy indicating whether there is at least one co-author in the same institution of the referee—, Same City—the fraction of the authors that are in an institution in the same city as the referee—-, and Same City (Max.)—a dummy indicating whether there is at least one co-author in an institution in the same city of the referee), respectively. Standard errors clustered by paper and referee in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: The Impact of the Authors' Gender on the Probability of Acceptance: Mechanisms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sh. Male Authors	0.002						0.016	0.017	0.016
	(0.021)						(0.020)	(0.017)	(0.019)
Sh. Male Auth. x Male Referee	0.041*								
	(0.024)								
Sh. Male Auth. | Well-known		-0.021	0.007	0.017	0.011	0.001			
		(0.023)	(0.032)	(0.028)	(0.026)	(0.048)			
Sh. Male Auth. | Lesser-known		0.046***	0.031**	0.039***	0.037***	0.033**			
		(0.014)	(0.015)	(0.013)	(0.013)	(0.014)			
Sh. Male Auth. x Masc. Field ($\geq$ p50)							0.031		
							(0.027)		
Sh. Male Auth. x Masc. Field ($\geq$ p75)								0.046	
								(0.029)	
Sh. Male Auth. x Macro*									0.018
									(0.032)
Sh. Male Auth. x Finance*									0.117**
									(0.049)
Sh. Male Auth. x Theory*									0.009
									(0.048)
Observations	16126	16126	16126	16125	16125	16126	16126	16126	16126
R^2	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
Well-known based on		Top-100 Ins.	Top-50 Ins.	No. Pub.	No. Pub. (10 Y)	No. Pub. (Top 5)			

Results of regressing a dummy indicating whether the paper was accepted on the indicated set of variables. In addition to the reported coefficients, all columns contain the full set of controls and fixed effects included in column (9) of Table 2. Columns (2)-(6) include number-of-well-known-authors dummies on top of number-of-authors dummies. Well-known and lesser-known authors defined in each column as specified on the last row of the table, for example, column (2) is based on Top-100 Ins., meaning that we code an author as well-known if he or she is in a top-100 institution. Masc. Field ($\geq$ p50) ($\geq$ p75) is a dummy indicating if the field is above the median (75th percentile) share of male authors. Fields above the median are econometrics, theory, finance, industrial organization, law and economics, macroeconomics, political economy, and public economics. Fields above the 75th percentile are theory, finance, macroeconomics, and political economy. Macro*, Finance*, and Theory* are dummies indicating the macro-area of the paper. Fields in empirical micro (the omitted category) are applied micro, behavioral, development, geography, history, environmental, industrial organization, labor, law and economics, and public. Fields in the macroeconomics macro-area are macroeconomics, international, and econometrics. Fields in the theory macro-area are theory and political economy. Standard errors clustered by paper and referee in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: The Impact of the Authors' Gender: Alternative Outcomes

	Accepted	Grade	Nomination	Accepted	Grade	Nomination
	(1)	(2)	(3)	(4)	(5)	(6)
Sh. Male Authors	0.0275	0.137*	0.0257*	0.0370*	0.119	0.0224
	(0.0195)	(0.0743)	(0.0150)	(0.0195)	(0.0757)	(0.0151)
Observations	5825	5825	1949	5646	5646	1924
R^2	0.258	0.316	0.229	0.261	0.320	0.232
Sample	All	All	All	No overrules	No overrules	No overrules

Results of regressing the indicated outcome on the share of male authors and the full set of controls and fixed effects included in column (9) of Table 2. All columns restrict the sample to observations in which Grade is non-missing. Columns (4)-(6) also drop observations in which the program committee overruled the recommendation of the referees. Standard errors clustered by paper and referee in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Online Appendices

Dataset Construction

The procedure to build the database takes place in several stages. Each year, submissions are identified by a unique *paper id*. To guarantee confidentiality, different elements of the files have been processed separately in order to codify the following variables:

Gender variables. We temporarily stored authors' and referees' names in a different file in order to codify gender, and deleted names afterwards. Gender is inferred from the first names of the authors using three different packages used in previous literature: (i) the R-package gender, constructed from U.S. Social Security data, (ii) the database constructed by Tang, Ross, Saxena, and Chen (2011) from Facebook lists of names, and (iii) a list constructed by Bagues and Campa (2017) with data from the Spanish National Institute of Statistics (INE). To identify a name as male or female, we imposed a minimum probability that a name is of a given gender of 80%. Names that could not be identified with the previous procedures, or for which we obtained conflicting genders, were completed by hand by a research assistant, inferring gender from photos or pronouns on the webpages or CVs of the authors. Papers for which in the end we cannot identify the gender of the author or, for multiple-authored papers, of all the co-authors, are dropped from the analysis (2.1% of observations).

Fields variables. For all the SMYE submissions, and 65% of SAEe submissions, we have a variable showing the field indicated by the authors at the time of submission. In these cases, we coded our fifteen field variables from this information. For the remaining cases (EEA and 35% of SAEe submissions), we temporarily stored the papers' titles on a different file and assigned the field by hand from the titles. When the field was not obvious from the title, we used JEL codes (available for years 2016 and 2017 of the EEA) and the abstract of the paper (which we searched for online). Two observations in which, after this procedure, we could not identify the field, are dropped from the analysis. After this, we deleted the file with the titles.

Cites variables. We temporarily stored the title of each paper in a separate file. We instructed our research assistants to look for each paper in Google Scholar, and to collect the number of cites at the time of the search (March 2019) and at one year before the paper was submitted (by restricting the search to papers dated until that year). After this procedure, we deleted the titles.

Number of publications variables. We temporarily stored the names of the authors in a different file. To obtain the prior record publication of the authors, we downloaded from Econlit the names of all the authors who have published at any of the journals listed in Table A2 since 2002. We then merged those names with the names in our dataset. After this merge, we deleted names.

Institutions variables. We temporarily stored the authors' institutions in a different file, and matched those institutions with their position at the ranking of *IDEAS/RePEc*, as of 3 December 2018, or with the 2018 QS World University Ranking. After this merge, we deleted institution names.

Same institution and city variables. We temporarily stored authors' and referees' institutions, and coded perfect matches as 1. Then we went over the remaining observations, and coded the variables Same Institution and Same City by hand. After this, we deleted institution names.

Shortest path variables. In the original temporal file that only included authors' and referees' names, we added the Repec identifiers for authors and referees and deleted names. Then we created a file with the codes of each pair author-referee in our sample, and scraped the data from collec.repec.org (accessed on October 2019) to obtain the shortest path for each pair author-referee. Finally, we merged this back to our main dataset using the identifiers.

Appendix Tables

Table A1: Summary Statistics, Mean of Variables by the Gender of the Authors

Sample	All	Maj. Male Authors	Half-Male and Half-Female	Maj. Female Authors
Accepted	0.52	0.54	0.52	0.47
Grade	5.73	5.76	5.83	5.58
Nomination	0.07	0.07	0.07	0.05
Sh. Male Authors	0.69	0.96	0.50	0.04
Half Male Authors	0.12	0.00	1.00	0.00
Majority Male Authors	0.65	1.00	0.00	0.00
Majority Female Authors	0.24	0.00	0.00	1.00
Male Referee	0.76	0.78	0.77	0.72
Number of Authors	1.77	1.82	2.12	1.48
Applied Micro	0.14	0.11	0.19	0.20
Behavioral	0.06	0.05	0.06	0.06
Development	0.05	0.04	0.05	0.08
Econometrics	0.04	0.04	0.03	0.03
History	0.01	0.01	0.01	0.01
Theory	0.06	0.07	0.05	0.04
Environmental	0.01	0.01	0.02	0.02
Finance	0.10	0.11	0.10	0.08
IO	0.05	0.06	0.03	0.05
International	0.07	0.07	0.08	0.08
Labor	0.06	0.05	0.07	0.08
Law and Economics	0.00	0.00	0.00	0.00
Macroeconomics	0.20	0.22	0.18	0.15
Political Economy	0.03	0.04	0.02	0.03
Public	0.10	0.11	0.11	0.09
Cites	0.13	0.15	0.13	0.10
Cites Ex Post	0.99	1.05	1.01	0.82
Published	0.08	0.09	0.08	0.07
No. Publications	0.36	0.43	0.43	0.15
No. Publications (10 Y)	0.62	0.72	0.76	0.25
No. Publications (Top 5)	0.06	0.07	0.06	0.03
Top 50 Ins.	0.11	0.11	0.07	0.11
Top 50-100 Ins.	0.11	0.11	0.11	0.11
Top 100-200 Ins.	0.19	0.20	0.19	0.16
Top 200-500 Ins.	0.34	0.32	0.39	0.34
Top 500-10% Ins.	0.09	0.09	0.12	0.08
Below 10% Ins.	0.17	0.17	0.11	0.19
Experience ≤ 0	0.27	0.25	0.13	0.37
Experience $\in (0,5]$	0.26	0.26	0.29	0.26
Experience $\in (5,10]$	0.20	0.19	0.29	0.16
Experience $\in (10,15]$	0.12	0.12	0.16	0.07
Experience $\in (15,20]$	0.06	0.07	0.07	0.04
Experience > 20	0.04	0.05	0.04	0.02
Experience Missing	0.05	0.05	0.03	0.07
Shortest Path	4.07	4.07	3.95	4.17
Shortest Path Missing	0.44	0.41	0.30	0.57
Same Ins.	0.01	0.01	0.01	0.01
Same City	0.02	0.02	0.02	0.02
Observations	16126	10408	1888	3830

The unit of observation is a pair paper-referee. The variable Grade is available only for the SMYE and for the EEA in 2016 and 2017, and the variable Nomination is available only for the SMYE.

Table A2: 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	

This list of 35 journals is used to create our publications variables and has been obtained from Card, DellaVigna, Funk, and Iriberri (2019).