Gender and equality at top economics journals*

Erin Hengel[†] Eunyoung Moon[‡]

November 2022
(First version: January 2018)

Abstract

Using (asinh) citations as a proxy for quality, we show that female-authored papers published in top economics journals are, on average, higher quality than male-authored papers. Furthermore, men and women publish higher quality papers when they co-author with women instead of men—for example, the same senior male economist receives about 60 log points more citations when he co-authors with a junior woman as opposed to a junior man. Finally, variance in quality is consistently higher among published male-authored papers. Under strong—but we believe reasonable—assumptions, we argue that these findings imply top economics journals hold female-authored papers to higher standards and, consequently, do not publish the highest quality research. Our results also suggest that popular proxies of academic impact discount women's contributions and that existing co-authoring relationships in economics under-exploit the capacity of female researchers.

Keywords: Gender, Discrimination, Quality, Citations, Research, Productivity, Collaboration; *JEL*: A11, J16, J24.

*This is a revised version of an earlier paper titled "Gender differences in citations at top economics journals" (January 2018). We especially thank Lunna Ai for excellent research assistance, Olga Gorelkina for help with the proof of Theorem 3.1 and Marco Scarsini for suggesting several references. We are also grateful for valuable feedback from Stefano DellaVigna, Jeremy Edwards, Guido Imbens, Balázs Muraközy, Stephen Ross, Anne Winkler, audience members at the 2019 American Economic Association and European Economic Association Annual Meetings and seminar participants at the Universities of Liverpool, Nottingham, Portsmouth and Duisburg-Essen. All errors are definitely our own.

†London School of Economics, erin.hengel@gmail.com.

‡University of Liverpool, emoon@liverpool.ac.uk.

1 Introduction

Publications in "top-five" economics journals are heavily weighted in tenure, promotion and salary decisions (Gibson *et al.* 2014; Heckman and Moktan 2019). They also serve as a signal for quality and integrity to policy makers and the media. As a result, they probably have a disproportionate impact on who succeeds in the profession, the research they produce, and the general direction of economic policy.

Top-five journals do not, however, publish very many papers by female authors. According to our data, women make up only 11 percent of all authors published since 1990, 12 percent since 2000 and 14 percent since 2010. Even in 2015, the average share of female authors per paper was still 15 percent. Only seven percent were majority female-authored, and just four percent were written entirely by women. In several recent years, the *Quarterly Journal of Economics*, *Econometrica* and the *Journal of Political Economy* did not publish a single exclusively female-authored manuscript.

In this paper, we ask whether higher standards for female authors contribute to their under-representation in top-five journals. To study our question, we construct a database of bibliographic and demographic information for almost 11,000 full-length papers published between 1950–2015 in the *American Economic Review* (*AER*), *Econometrica* (*ECA*), *Journal of Political Economy* (*JPE*), *Quarterly Journal of Economics* (*QJE*) and *Review of Economic Studies* (*REStud*). To proxy for quality, we use citations and adjust for potential confounders—including time since publication, field and the Matthew effect—by transforming them with the inverse hyperbolic sine function (asinh) and controlling for co-author count, author seniority and reputation and journal-year and *JEL* fixed effects (primary, secondary and tertiary).[1] To correct for citation practices that may differ across field, we also control for the length of a manuscript's bibliography.

Because these data are a selected sample, our analysis is guided by a theoretical framework that makes assumptions about the (unobserved) distribution of quality among submissions. Our framework identifies three conditions to determine whether female-authored papers are held to higher standards if quality is normally (although not necessarily identically) distributed among an unobserved set of male- and female-authored submissions (Theorem 3.1): (1) the mean quality of accepted female-authored papers is higher than the mean quality of accepted male-authored papers; (2) the variance in quality of accepted female-authored papers is no larger than the variance in quality of accepted male-authored papers; and (3) the mean acceptance rate for male-authored papers is the same as the mean acceptance rate for female-authored papers.

When we define the (unobserved) set of submissions to be the population of full-length manuscripts submitted to top-five journals, our evidence suggests that Theorem 3.1's three conditions hold. We find that accepted female-authored papers receive, on average, 8–9 log points more citations compared to male-authored papers (Condition 1); that rises to 16 log points after adjusting for the Matthew effect. Conclusions are roughly similar when estimated without these controls on a sample of papers likely less affected by the Matthew effect—*i.e.*, papers published after 2000 (for a discussion, see Section 4.1); conditional on field, however, the coefficient on female is only significant once the Matthew effect is taken into account.

Meanwhile, variance in quality is consistently higher among male-authored papers than it is among female-authored papers, conditional on acceptance (Condition 2).[2] Although we lack the data to test Condition 3, evidence from other studies suggests that male- and female-authored submissions to other general interest economics journals are accepted at similar rates (see, *e.g.*, Card *et al.* 2020).

[1] According to the Matthew effect, "winners" (*e.g.*, of prestigious awards) experience an artificial jump in status compared to otherwise identical "losers" (Merton 1968).

[2] As we discuss in Section 5.1, however, higher mean quality among female-authored papers combined with higher variance among male-authored papers (conditional on publication) is *not* consistent with the "greater male variability" hypothesis.

We replicate these results using several alternative ways to capture a paper's gender composition, proxy for quality using the log of 1 plus citations and test Condition 1 using raw counts as the dependent variable in negative binomial and quantile regression models. We also control for the length of an article's reference list, authors' institutions and non-parametrically account for co-author counts. In all instances, our evidence suggests that female-authored submissions to top-five journals are held to higher standards than are male-authored submissions.

We next define the set of submissions as the population of co-authored papers submitted to top-five journals by a single individual. Controlling for author and journal-year fixed effects, we find that men's accepted co-authored papers receive 11 log points more citations when they are co-authored with at least one woman; conversely, female authors receive 13–34 log points *fewer* citations when they are co-authored with at least one man. These results do not dramatically change after accounting for the Matthew effect or *JEL* fixed effects. Among male authors, however, the gap is somewhat sensitive to controlling for number of co-authors, which could be evidence that male authors are more likely to collaborate with high-quality men on projects with at least one female co-author.

To investigate, we restrict our sample to senior male economists with at least two top-five papers co-authored with a single junior author of each sex. This creates a treatment group—senior male authors co-authoring with exactly one junior woman—that very closely resembles the counterfactual group—those very same seniors co-authoring with exactly one junior man.

Controlling for author and journal-year fixed effects, we find that senior men's papers receive 21 log points more citations when they are co-authored with junior women as opposed to junior men, although the gap is not statistically significant. Once the Matthew effect and field are taken into account, however, it almost triples and becomes highly significant. We therefore conclude that accepted papers by senior men are higher quality when they are co-authored with junior women (as opposed to junior men), and contributions from unobserved co-authors, if anything, bias downward our estimates of the gender quality gap in multi-authored papers.

Finally, variance in quality is consistently lower when men *and* women co-author with women (Condition 2), and evidence in Card *et al.* (2018, p. 2018) suggests no statistically significant difference in acceptance rates between papers co-authored by one or more women compared to papers co-authored entirely by men (Condition 3). As before, we also replicate our results using the log of 1 plus citations and raw counts as a proxy for quality and control for the length of an article's reference list, authors' institutions and fixed effects for number of co-authors. We always find that men's and women's papers are higher quality when they are co-authored with women instead of men.

Combined, our evidence suggests journals subject female authors to higher standards and, as a result, their articles are better quality, conditional on acceptance. Nevertheless, there are several reasons to be cautious before coming to this conclusion. First, when the set of submissions is defined as the population of co-authored papers by a single individual, Theorem 3.1's three conditions must all hold simultaneously for that same author; however, we just test if they are satisfied on average. These particular results should therefore be interpreted as providing suggestive evidence of higher standards, only.

Second—and as emphasised in Theorem 3.1—higher standards only apply if (i) transformed citations are not biased in woman's favour, conditional on quality, and (ii) quality is normally (although not necessarily identically) distributed in the relevant populations of male- and female-authored submissions. We discuss the former assumption in detail in Section 4.1. Briefly, however, a large body of research consistently finds that female authors are more likely to cite female-authored papers than male authors are (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Ferber 1988) even among very similar manuscripts (Koffi 2019); as a result, we believe our estimates represent lower bounds on gender differences in quality at the mean.

Furthermore, it seems reasonable to suppose that most submissions to top-five journals are of roughly similar quality, very few are really good or really bad and the distribution is symmetric about the mean; thus, we consider normality to be a credible assumption when the set of submissions is the population of full-length manuscripts. In our opinion, normality is more plausibly violated among the population of co-authored papers by a single individual—for example, senior male authors might only submit their co-authored manuscripts to top-five journals when quality exceeds a threshold that many fail to meet. As we discuss in Section 4.2.2, our results in this case would still be informative about the presence of higher standards, just not about who, precisely, is responsible for setting them—*i.e.*, it could be editors and/or referees applying higher acceptance standards or the authors themselves applying higher co-authoring or submission standards.

Journals function as price mechanisms—*i.e.*, the journals in which articles are published serve as nominal currency for their value. If women could hedge (without friction) against every possible publication outcome in every possible state of the world, then biased acceptance decisions at one journal could simply be "undone" by a costless change in one's submission and publication strategy the previous date—*e.g.*, women could simply publish their higher quality papers in currently lower-tiered journals, confident that their actions would lead to an appropriate relative change in journal rankings the very next period.

When competition isn't perfect, however, discrimination interacts with one or more market frictions to prevent those who discriminate from fully internalising its costs. Consequently, its victims will have to *partially* bear them. For example, imperfect information about journal rankings may mean tenure and promotion committees' expectations are slow to adjust to the lower quality of journals that reject too many women.[3] As a result, women (and the men they co-author with) are tenured and promoted at lower rates than they otherwise would be if markets were complete and perfect. To the extent that grant committees similarly rely on applicants' past publication histories to choose between projects, women will also have a harder time funding future work.

Moreover, discrimination undoubtedly distorts authors' decisions in ways that can further misallocate available resources. Indeed, our own evidence implies male and female economists are better off collaborating with men, all else equal. This creates an incentive for authors of both sexes to forgo higher quality co-authoring opportunities with women in order to partner with men (see also Knobloch-Westerwick *et al.* 2013).

This paper makes several contributions to the literature. First, we contribute to a substantial body of research suggesting women are, in many situations, subjected to tougher standards and/or evaluated differently than men (see, *e.g.*, Card *et al.* 2020; Foschi 1996; Hengel 2022; Hospido and Sanz 2021; Krawczyk and Smyk 2016; Moss-Racusin *et al.* 2012; Reuben *et al.* 2014; Sarsons *et al.* 2019). Most relevant to our paper, Card *et al.* (2020) study manuscript submissions to the *Journal of the European Economic Association*, *Review of Economics and Statistics*, *QJE* and *REStud*. They take a different approach and identification strategy but find results roughly in line with and complementary to our own, namely that exclusively female-authored submissions receive about a quarter more citations than observably similar male-authored submissions.[4] Nevertheless, there are important differences between our two papers. In particular, Card *et al.* (2020) perform a between-paper comparison and find that mixed-gendered papers with a senior male co-author are not cited more than papers co-authored entirely by men; in contrast, we come to the opposite conclusion using within-author comparisons (for further discussion, see Appendix D.9).

Our second contribution is to the growing literature questioning common definitions of "research quality"

[3]See Heckman and Moktan (2019) for evidence that tenure expectations are indeed sticky.

[4]See also Koffi (2021) for descriptive evidence of a similarly sized positive association between female authorship and citations to papers published in top-five economics journals.

and studying how they materially impact women's visibility and perceived academic productivity. Of particular relevance is Zacchia (2021). She shows that the rankings of women in popular "top economists" lists decline as the weight given to journal articles increases. We complement her work by documenting evidence that productivity proxies relying on top-five publication counts will likely underestimate the productivity of female economists relative to male economists. Combined, Zacchia (2021) and our results suggest that women's contributions may be (unintentionally) discounted in many popular proxies of academic impact.

Relatedly, we join an emerging literature studying how gender differences in "market power" when choosing collaborators affects women's productivity. Our results from analysing returns to co-authoring suggest that women are held to higher standards in co-authoring relationships; however, they do not identify the party responsible for setting those standards. One possible explanation is that (senior) men prefer co-authoring with other men and collaborate with women only when the expected value of their joint output is especially high. This interpretation is congruent with suggestive evidence in Gertsberg (2022). She finds that female economists' research output declined in the post-#MeToo era because male economists perceived a greater risk of being falsely accused of sexual harassment. Combined, Gertsberg (2022) and our results suggest that male economists may not fully internalise the negative externality their preference for co-authoring with other men has on women's productivity; as a result, existing co-authoring relationships in economics may under-exploit the capacity of female researchers.

Fourth, we also contribute to the methodological literature on outcome tests (see, *e.g.*, Anwar and Fang 2006; Arnold *et al.* 2018; Knowles *et al.* 2001; Marx 2021). Originally developed by Becker (Becker 1957; Becker 1993), outcome tests compare group measures of success conditional on outcome—*e.g.*, if there were no gender bias in peer review, then marginally accepted male- and female-authored papers should be the same quality. Unfortunately, however, marginal outcomes are usually unobserved, and average outcomes often poorly proxy for them (see, *e.g.*, Ayres and Waldfogel 2006; Simoiu *et al.* 2017). To overcome this "infra-marginality" problem, we develop a simple test that relies on distributional assumptions about male- and female-authored submissions. This allows us to identify conditions where the average quality of accepted papers is also informative about the quality of marginally accepted papers. We believe our test can provide useful policy-relevant information in cases where it is difficult or impossible to identify the marginal unit of assessment, and there are strong *a priori* grounds to believe that unconditional, group-specific measures of success are normally distributed.

Finally, our paper also contributes to the literature studying *why* female academics publish fewer papers than men (see, *e.g.*, Alexander *et al.* 2021; Ductor *et al.* 2021) and how that links to their persistent under-representation in economics and related disciplines (Bateman *et al.* 2021; Bayer and Rouse 2016; Chari and Goldsmith-Pinkham 2017; Gamage *et al.* 2020; Ginther and Kahn 2004; Lundberg and Stearns 2019; Teele and Thelen 2017).

Our paper proceeds in the following order. Section 2 discusses our data as well as the representation of women in top economics journals. Sections 3 and 4 introduce the theoretical and empirical strategies we use to identify higher standards. Section 5 presents the results, and Section 6 concludes.

2 Where are the women?

The dataset we analyse contains basic bibliographic information and data on author characteristics for 10,951 regular issue, full-length, original research articles published between 1950–2015 in the *AER*, *ECA*, *JPE*, *QJE* and *REStud*.[5] These data were originally collected and analysed in Hengel (2022); for

[5] We define "regular issue, full-length, original research" articles as any non-errata/corrigenda/editorial article published with an abstract, excluding *Papers & Proceedings* issues of the *AER*. Before 1980, our sample is disproportionately made up of articles published in *ECA*, *JPE* and *REStud*, which systematically published abstracts with their full-length, original

further details on sources, coverage, collection procedures and variable definitions, see Appendix B and Hengel (2022).

Our data suggest that female authors are very under-represented in top-five journals. As Graph (A) in Figure 1 illustrates, the situation has improved little with time: women make up only 11 percent of all authors published since 1990, 12 percent since 2000 and 14 percent since 2010.[6] Between 1986–2015, there has been zero growth in the number of exclusively female-authored papers; almost no growth in the number of majority female-authored papers; and no meaningful change in the number of mixed-gendered papers with a senior female co-author. The only tepid growth that *has* occurred, is largely—if not entirely—due to an increase in the number of articles by senior men co-authoring with a weak minority of junior women.

Top-five journals publish about as many solo female-authored papers today as they did in the late 1980s (Figure 1, graph (B)): seven in 1986, ten in 1997 and eleven in 2015. The number of solo male-authored papers, however, has declined: 125 were published in 1986, 62 in 1997 and 45 in 2015. As a result, the proportion of solo-authored papers by women has increased from five percent in 1986 to twenty percent in 2015.

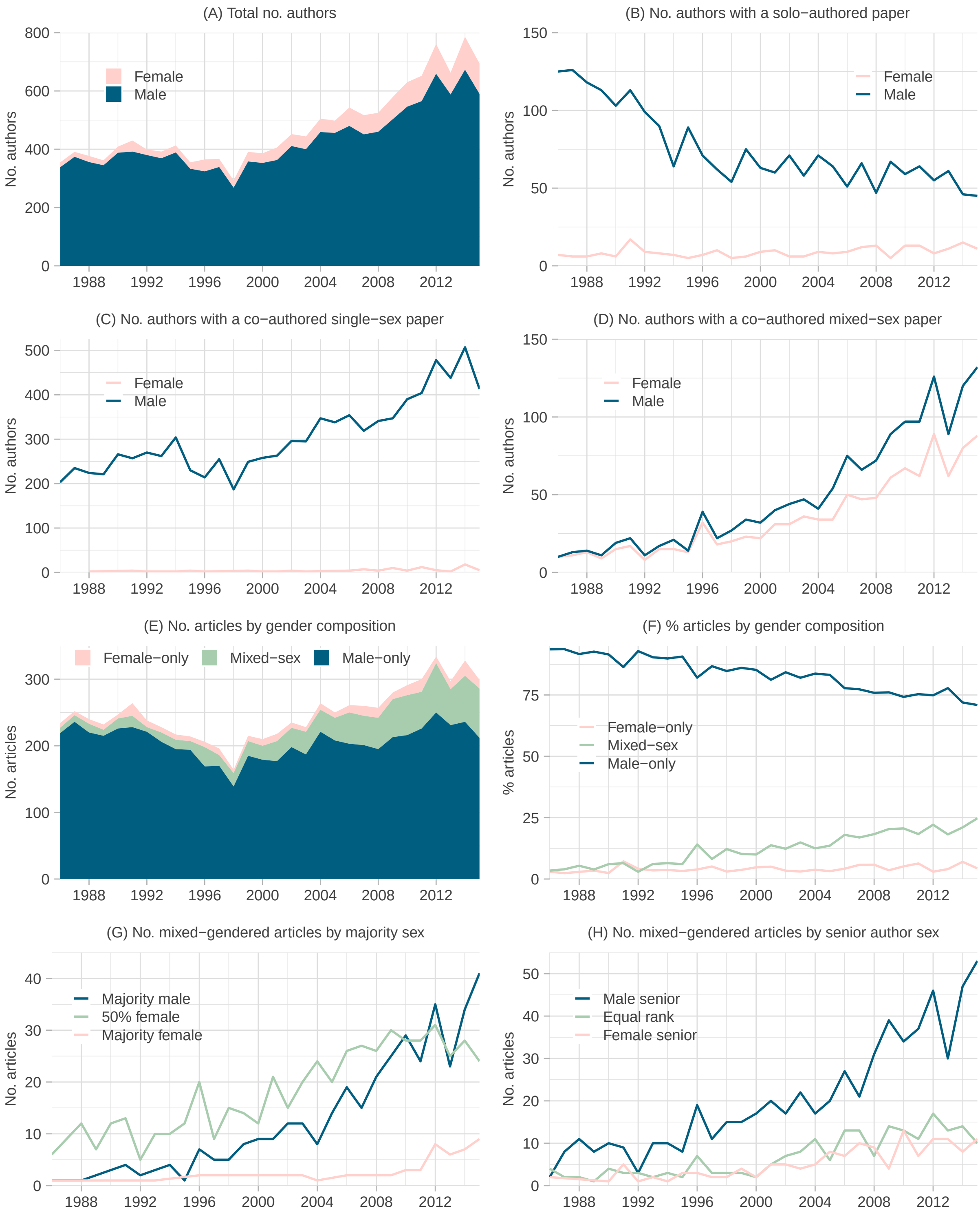
But falling male solo-authored papers has been more than offset by rising male *co-authored* papers. Consequently, the proportion of female authors on single-sex papers has remained stubbornly close to zero for the past 30 years (Figure 1, graph (C)). In 1987, top-five journals collectively published 96 articles co-authored by two men and zero articles co-authored by two women; in 2015, the corresponding figures were 102 and one. Meanwhile, journals have sharply increased the number of single-sex articles they publish by three or more men: 65 were published in 2015 versus 15 in 1986. As of 2015, however, only ***six*** had *ever* been published by women; no top-five journal had yet to publish a full-length paper exclusively authored by four or more women.

Moreover, women *do not* make up a greater share of authors on mixed-gendered papers. Journals are publishing more articles with at least one female author, but the number of male authors on these papers has increased slightly faster than the number of female authors—meaning the share of women among authors on mixed-gendered papers has actually *declined*. Graph (D) in Figure 1 plots the number of authors with a co-authored mixed-sex top-five paper each year. In the late 1980s, men and women were about equally represented. Since then, however, mixed-gendered papers have tended to generate more publications for men than they do for women. Graphs (E) and (F) reinforce this conclusion. They plot the number and percentage of single- and mixed-gendered papers published in top-five journals, respectively: the latter has increased, but the former has not.

Finally, majority- and senior-female-authored papers are almost as rare today as they were 30 years ago. Very few majority-female mixed-sex papers were published in top-five journals before 2000; since then, they publish about four a year (Figure 1, Graph (G)). Meanwhile, the number of mixed-gendered papers with a majority or equal share of *male* authors has risen. The result is little or no growth in majority-female papers. Similarly, mixed-gendered papers with male senior authors have steadily increased since the late 1980s (Graph (H)); growth in papers with a senior female author or male and female co-authors of equal rank, however, has not.

research articles before *AER* and *QJE*. Starting in the mid-1980s, however, almost all full-length original research papers in any top-five journal were published with an abstract.

[6] In contrast, women are somewhat better represented in university economics departments. For example, women were 26 percent of academic economists at UK universities in 2018—33 percent of lecturers, 27 percent of senior lecturers/readers and 15 percent of professors (Bateman and Hengel 2022). Figures from the US are roughly similar (Chevalier 2021; Lundberg and Stearns 2019), but are slightly higher (33 percent) in continental Europe (Auriol *et al.* 2022).



Note. Graph (A) displays the stacked total number of female (pink) and male (blue) authors published in a top-five journal each year. Graph (B) is the (non-unique) number of male and female economists with a solo-authored paper; Graphs (C) and (D) plot the corresponding number of authors with a co-authored single-sex paper and a co-authored mixed-gendered paper. Graphs (E) and (F) are the stacked total number and percentage, respectively, of exclusively female-authored, mixed-gendered (green) and exclusively male-authored papers. Graphs (G) and (H) plot the total number of mixed-gendered papers with a strict majority of male and female co-authors and a male and female senior author, respectively; papers with an equal number of each gender or two or more senior authors of the opposite gender shown in green.

Figure 1: Gender composition of top-five publications

3 Theoretical framework

In this section, we construct a theoretical framework to help us evaluate whether higher standards contribute to the under-representation of women documented in Figure 1. Suppose q_k is an indicator that perfectly captures the quality of papers in $\mathcal{G}$, where $\mathcal{G}$ can be partitioned into male- ($\mathcal{G}_M$) and female-authored ($\mathcal{G}_F$) subsets. Assume q_k is normally (although not necessarily identically) distributed in both $\mathcal{G}_M$ and $\mathcal{G}_F$ and papers in $\mathcal{G}_g$ are accepted for publication when q_k exceeds θ_g, $g \in \{M, F\}$.

When these assumptions hold, Theorem 3.1 identifies three conditions that, if satisfied, establish that papers in $\mathcal{G}_F$ are accepted less often than papers in $\mathcal{G}_M$, conditional on q_k. First, the mean acceptance rate for papers in $\mathcal{G}_M$ is the same as the mean acceptance rate for papers in $\mathcal{G}_F$. Second, the variance in the quality of papers in $\mathcal{G}_F$ is no larger than the variance in the quality of papers in $\mathcal{G}_M$, conditional on acceptance. And third, the mean quality of papers in $\mathcal{G}_F$ is strictly greater than the mean quality of papers in $\mathcal{G}_M$, again conditional on acceptance.

Theorem 3.1. *Let $\mathcal{G}$ denote a set of papers that can be partitioned into male- ($\mathcal{G}_M$) and female-authored ($\mathcal{G}_F$) subsets. Assume:*

Assumption 1. There exists an indicator q_k that perfectly captures the quality of papers in $\mathcal{G}$.

Assumption 2. Papers in $\mathcal{G}$ are accepted if $q_k > \theta_g$, where θ_g is some threshold specific to $\mathcal{G}_g$, $g \in \{M, F\}$.

Assumption 3. q_k in $\mathcal{G}_M$ and q_k in $\mathcal{G}_F$ are both normally (although not necessarily identically) distributed with mean μ_g and variance σ_g^2, $g \in \{M, F\}$.

When these assumptions plus the following three conditions are satisfied, then $\theta_F > \theta_M$.

Condition 1. Conditional on acceptance, the mean of q_k in $\mathcal{G}_F$ is strictly larger than the mean of q_k in $\mathcal{G}_M$: $\mu_F(\theta_F) > \mu_M(\theta_M)$, where $\mu_g(\theta_g)$ is the mean quality of accepted papers in $\mathcal{G}_g$, $g \in \{M, F\}$.

Condition 2. Conditional on acceptance, the variance of q_k in $\mathcal{G}_F$ is not larger than the variance of q_k in $\mathcal{G}_M$: $\sigma_F^2(\theta_F) \leq \sigma_M^2(\theta_M)$, where $\sigma_g^2(\theta_g)$ is the variance in the quality of accepted papers in $\mathcal{G}_g$, $g \in \{M, F\}$.

Condition 3. The average acceptance rate of papers in $\mathcal{G}_M$ is the same as the average acceptance rate of papers in $\mathcal{G}_F$: $\Phi_F(\theta_F) = \Phi_M(\theta_M)$, where Φ_g is the cumulative normal distribution of q_k in $\mathcal{G}_g$ for gender $g \in \{M, F\}$.

Theorem 3.1 is proved in Appendix A. To understand its rough intuition, suppose $\theta_M = \theta_F = \theta$ and the proportion of accepted papers in $\mathcal{G}_F$ is the same as it is in $\mathcal{G}_M$. Under these conditions, greater variability in $\mathcal{G}_M$ means that the average q_k for male-authored papers is further to the right of θ compared to the average q_k for female-authored papers, conditional on $q_k > \theta$. Or in other words, the average quality of accepted papers in $\mathcal{G}_M$ is higher than the average quality of accepted papers in $\mathcal{G}_F$. When it isn't, $\theta_F > \theta_M$.

Theorem 3.1 is only valid if Assumptions 1–3 hold. Assumption 1 is discussed in more detail in Section 4.1. Given knowledge of q_k, Assumption 2 simply implies that higher quality papers are more likely to be accepted. Assumption 3 depends on the definition of $\mathcal{G}$. It would be violated, for example, if most q_k in $\mathcal{G}$ clustered around an upper or lower limit; thus, one must take care to define $\mathcal{G}$ so that $q_k \sim \Phi_g$ is normally distributed for both $g = M$ and $g = F$.

$\mathcal{G}$'s definition is also crucial in two other ways. First, conclusions drawn from Theorem 3.1 only apply to papers in $\mathcal{G}$; as a result, they may not be relevant for other populations—*e.g.*, manuscripts submitted to journals not covered by $\mathcal{G}$. Second, the definition of $\mathcal{G}$ identifies the parties responsible for applying $\theta_F \neq \theta_M$. To see this, suppose $\mathcal{G}$ includes all manuscripts submitted to top-five journals. In this case

Theorem 3.1 would establish whether editors and/or referees hold female-authored submissions to higher standards. Alternatively, if $\mathcal{G}$ were defined as the set of all *potential* co-authored papers by individual i, then Theorem 3.1 determines if i expects higher standards from female collaborators as a condition of co-authorship.

4 Empirical implementation

4.1 Citations as a proxy for q_k

Theorem 3.1's first assumption requires that q_k exists and is known. Because "quality" is not well-defined, however, it cannot be perfectly measured; instead we use citations as an imperfect proxy. Although papers are cited for a variety of reasons—including to criticise and correct—most studies find they positively correlate with peer assessments of research quality (see *e.g.*, Aksnes and Taxt 2004; Oppenheim 1997; Rinia *et al.* 1998; van Raan 2006). As a result, bibliometricians generally agree that citations roughly quantify (albeit noisily) the value of a scholarly output to its relevant research community (for further (and deeper) discussions, see *e.g.*, Aksnes *et al.* 2019; D'Ippoliti 2021).

Unadjusted citations do, however, suffer from several forms of measurement error that may bias their estimates of gender differences in quality at the mean. First, older articles have had more time to accumulate citations and are also disproportionately male-authored. Second, men and women differ in the number of people they collaborate with, and higher co-author counts may artificially inflate citations relative to quality—*e.g.*, by increasing a paper's scope to accumulate self-citations.[7] And third, different fields have different citation practices and norms and also vary in terms of female representation.

A fourth source of non-classical measurement error is the so-called "Matthew effect" (Merton 1968)—*i.e.*, fame begets more fame: "For to every one who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away" (Matthew 25:29, Revised Standard Version). In the context of citations, the Matthew effect skews the distribution's right-tail beyond what is probably justified by differences in quality. Since men's papers are more prevalent at this end of the distribution (Figure 2, Graph (A)), the skew is likely greater for them than it is for women. As a result, citations arguably give too much weight to a small number of highly cited—and disproportionately male-authored—papers when used as a proxy for quality.[8]

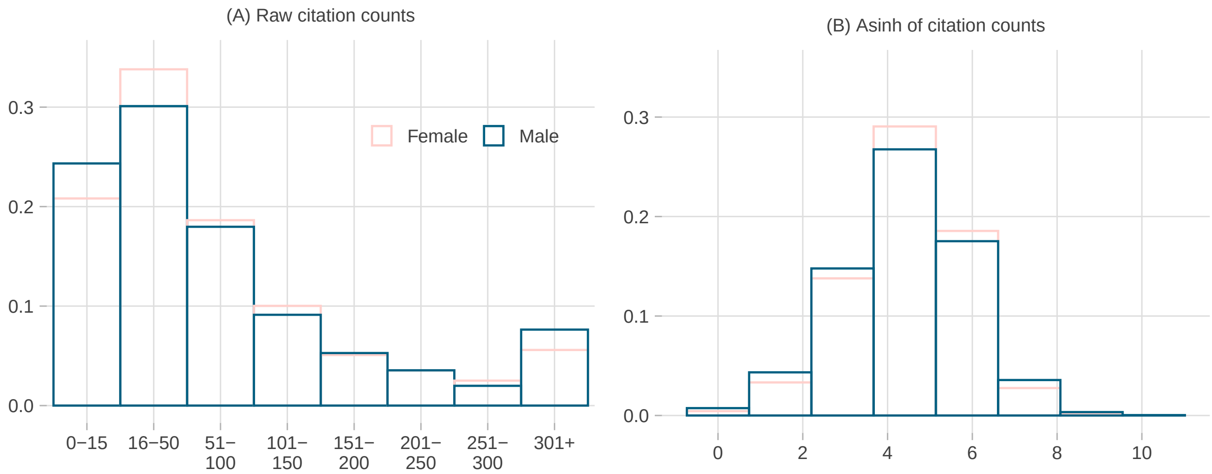
A final source of measurement error is bias against women in the decision to cite. A large body of research analyses manuscript bibliographies to determine whether female authors are more likely than male authors to cite female-authored papers—and consistently finds that they are (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Ferber 1988) even among very similar manuscripts (Koffi 2019). Thus, citation counts probably under-estimate the quality of female-authored papers—and over-estimate the quality of male-authored papers—across the entire distribution of citations.[9]

We account for the first two forms of measurement error—*i.e.*, time since publication and number of co-authors—by controlling for journal-year fixed effects and co-author counts. To adjust for field-specific citation practices and norms, we include fixed effects for primary, secondary and tertiary *JEL* categories. Because *JEL* codes are only as good as the legitimacy and accuracy of the *JEL* classification system, we also apply an alternative approach common in the bibliometric literature: citing-side normalisation.

[7] For a variety of reasons, co-authored papers may also be higher quality (see, *e.g.*, Ahmadpoor and Jones 2019). We therefore always show results with and without controlling for the number of authors on a paper.

[8] For evidence of the "Matthew effect" in citations, see Azoulay *et al.* (2014). In Appendix E, we illustrate the impact it likely has on gender differences in mean raw citation counts by constructing and controlling for a set of "superstar" and Nobel Prize fixed effects.

[9] A related issue is that men have denser research networks (Ductor *et al.* 2021); as a result, male-authored research may be cited more, simply because male authors have stronger social ties to their colleagues (D'Ippoliti 2021; D'Ippoliti *et al.* 2021).



Note. Left-hand graph displays the fraction of authors with a top-five paper that was cited 0–15 times, 16–50 times, 51–100 times, *etc.* Right-hand graph plots the histogram of transformed citations (asinh).

Figure 2: Distribution of citations

Evidence suggests that differences between fields in citation densities are largely driven by field-specific differences in the propensity to cite (for further discussion and the related literature, see Waltman 2016); on way to correct for this is by adjusting the for the length of a manuscript's reference list.

To temper the Matthew effect, we adjust for both the *immediate* impact of co-author fame—measured as the most prolific co-author's total number of top-five articles at the time a paper was published ($\max t$)—as well as the *delayed* effect[10]—measured as the most prolific co-author's total number of top-five articles when citations were collected ($\max T$). For robustness, we also limit our sample to papers published after 2000; assuming early citations to an article are less susceptible to distortions caused by the Matthew effect (Aksnes 2003; Aksnes *et al.* 2019), this allows us to omit the partially endogenous controls $\max t$ and $\max T$.

Finally, often-cited papers are probably cited more, conditional on quality, even after accounting for $\max t$ and $\max T$. We therefore also transform raw citation counts with the inverse hyperbolic sine function (asinh); this reduces the impact of outlier observations while preserving rank order (Figure 2, Graph (B)). We do not, however, explicitly adjust for general bias against women in the decision to cite. As a result, even asinh citations probably under-estimate the quality of women's research relative to men's. Thus, our results likely represent lower bounds on gender differences in quality, conditional on $\mathcal{G}$.

4.2 Estimation strategy

4.2.1 Submissions to top-five journals

Suppose $\mathcal{G}$ is the set of all papers submitted to top-five journals. To determine whether Theorem 3.1's Conditions 1 and 2 are satisfied, estimate Equation (1) using data on accepted papers in $\mathcal{G}$:

$$\hat{q}_k = \beta_0 + \beta_1 \,\text{female}_k + \beta_2 \, N_k + \beta_3 \, \max t_k + \beta_4 \, \max T_k + \theta \, \mathbf{X}_k + \varepsilon_k, \tag{1}$$

where $\hat{q}_k$ is our proxy for quality (asinh citations), female_k an indicator equal to 1 if paper k is female-authored, N_k the number of co-authors on k, $\max t_k$ the seniority of its most senior co-author, $\max T_k$ the prominence of its most prominent co-author, $\mathbf{X}_k$ a vector of journal, year and *JEL* fixed effects, and ε_k the error term. (Our reasons for including these variables are discussed in Section 4.1.)

[10] That is, the citations a paper accumulates aren't fixed in time. As a result, they could be influenced by the future success or failure of a paper's authors. Thus, a stronger publishing record later on probably drives citations to earlier work, all else equal (see, *e.g.*, Azoulay *et al.* 2014; Bjarnason and Sigfusdottir 2002).

If Assumptions 1–3 in Theorem 3.1 hold, then the sign and significance of β_1 in Equation (1) determines whether Condition 1 is satisfied.[11] For the second condition, separately estimate Equation (1) on male- and female-authored subsets to obtain the gender-specific variance of $\hat{q}_k$, conditional on acceptance. Condition 3 requires editorial outcomes for all papers in $\mathcal{G}$, which we do not have; however, evidence from other studies suggests men's and women's papers are accepted at roughly similar rates (for data specific to economics, see *e.g.*, Blank 1991; Card *et al.* 2020).

As discussed in Section 3, the definition of $\mathcal{G}$ determines which parties are potentially responsible for applying $\theta_F \neq \theta_M$. Suppose, for example, that Theorem 3.1 establishes that women are held to higher standards. Together, Assumptions 2 and 3 imply that both men and women submitted papers with $\hat{q}_k \in [\theta_M, \theta_F)$, but among them, only the male-authored submissions were accepted. Since editors and/or referees make these decisions, they set $\theta_F > \theta_M$.

4.2.2 Co-authored submissions to top-five journals by inidividual i

Now define $\mathcal{G}_i$ as the set of all co-authored papers by individual i that i also submits to top-five journals. To apply Theorem 3.1, estimate Equation (2) using data on accepted papers in $\mathcal{G}_i$:

$$\hat{q}_{it} = \alpha_i + \beta_1\, g_{it}^{-i} + \beta_2\, N_{it} + \beta_3\, \max t_{it} + \beta_4\, \max T_{it} + \theta\, \mathbf{X}_{it} + \varepsilon_{it}, \tag{2}$$

where α_i is an individual fixed effect and $g_{it}^{-i} \in \{F, M\}$ an indicator equal to 1 if i's tth top-five paper is co-authored with a member of the opposite sex (*i.e.*, $g_{it}^{-i} = \text{female}_{it}$ if i is male and $g_{it}^{-i} = \text{male}_{it}$ if she is female); $\hat{q}_{it}$, N_{it}, $\max t_{it}$, $\max T_{it}$, $\mathbf{X}_{it}$ and ε_{it} are author-level analogues of the variables defined in Equation (1).

As in Section 4.2.1, the sign and significance of β_1 in Equation (2) indicates whether Theorem 3.1's Condition 1 is satisfied; to obtain the gender-specific variance of $\hat{q}_{it}$ (Condition 2), separately estimate Equation (2) on i's accepted papers with male and female co-authors. Condition 3 requires data we do not have—*i.e.*, editorial outcomes for all of i's co-authored submissions. Unfortunately, we are also not aware of research specifically investigating whether an individual's acceptance rates differ when he co-authors with men vs. women. Nevertheless, there does not appear to be a statistically significant difference in acceptance rates among all co-authored papers by one or more women compared to all co-authored papers only by men (see Card *et al.* 2018, p. 280).

For several reasons, we encourage additional caution when applying Theorem 3.1 to $\mathcal{G}_i$. First, Conditions 1–3 must be satisfied for the same i. Thus, Equation (2) is ideally estimated on data from a single individual; when it isn't, conclusions drawn from it should be interpreted as suggestive, only.

A second issue relates to the distribution of quality across all i. Suppose that Theorem 3.1 establishes $\theta_{iF} > \theta_{iM}$. Editors and/or referees could have increased the quality of the papers they publish by accepting a greater fraction of i's co-authored papers with women. However, without making further distributional assumptions about the quality of *all* marginally rejected co-authored papers with women, we cannot conclude that accepting a greater fraction of them would also increase quality.

Finally, Theorem 3.1 would be violated if i submits papers to top-five journals only when their quality exceeds a threshold that many fail to meet. When this happens, $\mathcal{G}_i$ should be redefined to cover a population of i's papers that *is* normally distributed, although (and as discussed in Section 3) it may become more difficult to identify who decides $\theta_{iF} \neq \theta_{iM}$. For example, suppose $\mathcal{G}_i$ were redefined to include all i's co-authored papers (wherever they were submitted), but only citations to his top-five papers

[11] The most controversial of these assumptions is normality (Assumption 3), which effectively requires that most submissions are of roughly similar quality, very few are either really good or really bad and the distribution is symmetric about the mean. (Even if authors only submit their very best papers, Assumption 3 would still hold as long as the distribution of quality across all "very best papers" is itself normal.)

are observed. Unless we also explicitly assume that i's marginally accepted top-five paper was (or or was not) submitted to top-five journals, we cannot identify the exact party responsible for $\theta_{iF} \neq \theta_{iM}$—it could be editors, referees, i or all three.

5 Results

5.1 Submissions to top-five journals

Consider the case when the unobserved $\mathcal{G}$ is the set of all papers submitted to top-five journals. Table 1 displays results from OLS estimation of Equation (1) on the sample of accepted papers in $\mathcal{G}$. To determine the gender of paper k, we set $\text{female}_k = 0$ if all of its authors are male, $\text{female}_k = 1$ if at least 50 percent are female, and drop mixed-gendered papers that satisfy neither condition.

We first test Condition 1 of Theorem 3.1. Our evidence consistently suggests that the mean quality of female-authored papers exceeds the mean quality of male-authored papers, conditional on acceptance. Results in columns (1) and (2) suggest that female-authored papers receive on average 8–9 log points more citations; that rises to 17 log points after adjusting for the Matthew effect with $\max t$ (author seniority at the time of publication) and $\max T$ (author prominence at the time citations were collected).[12] Conclusions are roughly similar when Equation (1) is re-estimated on the sample of papers published after 2000 and without controlling for $\max t$, $\max T$ or N (see Section 4.1 for a discussion); conditional on primary *JEL* fixed effects, however, the coefficient on female is only significant once the Matthew effect is taken into account (columns (6)–(9)).

To assess the sensitivity of β_1 to omitted variables, we use information from selection on observables to bound potential bias from selection on unobservables (Altonji *et al.* 2005; Oster 2019). Table 1's third horizontal pane reports these bounds corresponding to the assumption that the unobservables explain about as much of the variation in the dependent variable as the observables do. According to column (9), they suggest that female-authored papers published in top-five journals receive about 12–28 log points more citations.

When we test Condition 2, we find that variance in quality is higher among male-authored papers than it is among female-authored papers, conditional on acceptance. Figure 3 plots the distribution of residualised asinh (right) and raw citations (left) among solo-authored manuscripts. Women's papers are relatively absent from the right- and (especially) left-hand tail of both distributions, suggesting a smaller variance compared to men's. Estimates of the variance of ε_k in male- and female-authored sub-samples confirm this. They consistently suggest that $\sigma_F^2(\theta_F)$ is smaller than $\sigma_M^2(\theta_M)$.

As for Condition 3, we lack the data to test whether male- and female-authored submissions are accepted at similar rates. Nevertheless, evidence from other studies suggests that they are (see Section 4.2.1 for a discussion). We therefore conclude from Theorem 3.1 that $\theta_F > \theta_M$.

For robustness, we replicate Table 1 using several alternative ways to capture a paper's gender composition (Appendix C.4) and proxy for q_k using the log of 1 plus citations (Appendix C.3).[13] We also re-estimate Equation (1) using raw counts as the dependent variable in negative binomial and quantile regression models (Appendices C.2 and C.1). In Appendix C.7 we control for secondary and tertiary *JEL* codes; in Appendices C.5 and C.8 we non-parametrically account for number of co-authors and control for the length of an article's bibliography, respectively; in Appendix C.6, we control for fixed effects for authors' institutional rank. In all instances, results are consistent with those reported in Table 1.

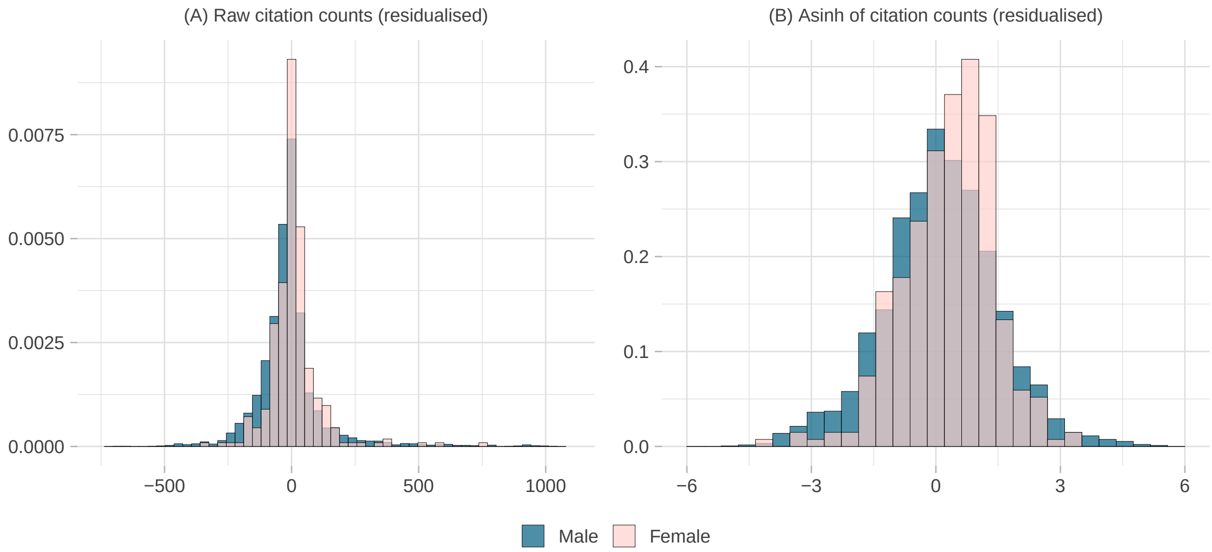
[12] To conserve space, we do not control for $\max t$ and $\max T$ separately, but these results are available in an earlier version of this paper (see Hengel and Moon 2020).

[13] Since Theorem 3.1 only applies when the gender of a paper is defined in binary terms, we do not replicate Table 1 using a categorical variable to account for different gender compositions on papers. These results are, however, available in an earlier version of this paper (see Hengel and Moon 2020, Appendix D Table D.1).

Table 1: Gender differences in the quality of published top-five papers

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female (β_1)	0.077**	0.092**	0.166***	0.093**	0.115***	0.173***	0.031	0.053	0.116***	0.141***	0.184***
	(0.038)	(0.038)	(0.037)	(0.039)	(0.039)	(0.038)	(0.039)	(0.039)	(0.038)	(0.043)	(0.044)
N		0.218***	0.173***		0.198***	0.166***		0.194***	0.161***		
		(0.017)	(0.017)		(0.018)	(0.018)		(0.018)	(0.018)		
$\max t$			−0.049***			−0.044***			−0.043***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.747	1.725	1.633	1.216	1.193	1.139	1.160	1.139	1.085	1.034	1.117
$\sigma^2_F(\theta_F)$	0.912	0.894	0.866	0.881	0.860	0.833	0.845	0.827	0.798	0.833	0.996
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.067
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
R^2	0.265	0.275	0.312	0.349	0.362	0.390	0.376	0.388	0.417	0.403	0.355
Bounds (β_1)	[0.03,0.08]	[0.06,0.09]	[0.17,0.21]	[0.09,0.22]	[0.11,0.26]	[0.17,0.39]	[0.03,0.10]	[0.05,0.15]	[0.12,0.28]	[0.14,0.20]	[0.18,0.28]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Figures correspond to coefficients from OLS estimation of Equation (1). The dependent variable is citation counts (asinh). Independent variables include: (i) a binary variable (female) equal to one if at least 50 percent of authors are female and 0 if they are all male (mixed-gendered papers with fewer than 50 percent female authors are dropped); (ii) number of co-authors (N); (iii) author seniority at the time of publication ($\max t$); and (iv) author prominence at the time citations were collected ($\max T$). $\sigma^2_M(\theta_M)$ and $\sigma^2_F(\theta_F)$ are residual variances from estimating Equation (1) in the samples of papers satisfying female = 0 and female = 1, respectively. They are followed by p-values from testing the null hypothesis $\sigma^2_M(\theta_M)/\sigma^2_F(\theta_F) = 1$. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.



Note. Graphs display the histograms of raw (left) and asinh transformed (right) citations for solo-authored papers. Citations have been residualised with respect to $\max t$, $\max T$ and journal-year fixed effects.

Figure 3: Distribution of citations (residualised), solo-authored papers

We end by noting that Table 1 and Figure 3 present evidence that is not consistent with the "greater male variability" hypothesis. Gender differences in variability are equivalent to gender differences in conditional averages. Presumably, academic economists—and especially those publishing in the best journals—are drawn from the top half of the distribution of "talent". Thus, greater variability among men implies that the average quality of male-authored papers is higher than the average quality of female-authored papers, conditional on publication in a top-five journal. Our evidence suggests that the opposite is true. It may also indicate that referees and editors are less willing to gamble on women's riskiest work. (See also Ball *et al.* (2020) for similar arguments and evidence using citation data from fundamental physics.)

5.2 Co-authored submissions to top-five journals by inidividual i

Now define $\mathcal{G}_i$ (which is unobserved) as the set of all co-authored papers by individual i that i also submits to top-five journals. In order to follow i over the $t \in \{1, \ldots, T_i\}$ co-authored papers he publishes in these journals, we duplicate each article N_k times and assign observation k_n article k's nth$\in \{1, \ldots, N_k\}$ co-author. We use the resulting panel dataset to estimate Equation (2) in an author-level fixed effects model. To determine the gender of i's co-authored papers, we set $g_{it}^{-i} = \text{female}_{it} = 1$ if i is male and his tth paper is co-authored with at least one woman; similarly, $g_{it}^{-i} = \text{male}_{it} = 1$ if i is a woman and her tth paper is co-authored with at least one man. (Solo-authored papers are dropped.)

The first panel of Table 2—which is estimated on the sample of male authors only—suggests that men's papers are higher quality when they are co-authored with women (Theorem 3.1, Condition 1). The coefficient on female is consistently positive and generally statistically significant (Theorem 3.1, Condition 1). According to column (1), men's papers receive 11 log points more citations when they are co-authored with at least one woman; the gap is roughly similar conditional on $\max t$ and $\max T$ (column (2)), but is somewhat sensitive to controlling for N (column (3)). Columns (4)–(9) suggest a similar pattern when the sample is restricted to papers published after 1990 and conditional on primary *JEL* fixed effects. In the final two columns, we re-estimate Equation (2) on papers published after 2000 and omit controls for N, $\max t$ and $\max T$ (see Section 4.1 for a discussion); coefficients and standard errors roughly resemble those reported in columns (2) and (5).

The sensitivity of β_1 with respect to N may indicate that it is biased by contributions from unobserved

Table 2: Returns to co-authoring with the opposite sex

	1990–2015									2000–2015	
	All data			without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	0.110***	0.125***	0.080**	0.117***	0.130***	0.070*	0.088**	0.101***	0.043	0.143***	0.150***
	(0.038)	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)	(0.039)	(0.040)	(0.041)
$\max t$		−0.010**	−0.012***		−0.013***	−0.016***		−0.014***	−0.016***		
		(0.004)	(0.004)		(0.004)	(0.004)		(0.004)	(0.004)		
$\max T$		0.020***	0.019***		0.023***	0.021***		0.023***	0.022***		
		(0.003)	(0.003)		(0.004)	(0.004)		(0.004)	(0.004)		
N			0.117***			0.143***			0.139***		
			(0.020)			(0.020)			(0.019)		
$\sigma^2_M(\theta_M)$	0.544	0.540	0.538	0.430	0.426	0.422	0.419	0.414	0.411	0.342	0.364
$\sigma^2_F(\theta_F)$	0.129	0.129	0.127	0.127	0.125	0.124	0.118	0.117	0.116	0.133	0.178
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−0.132	−0.235*	−0.314**	−0.191	−0.268**	−0.338**	−0.142	−0.225*	−0.324**	−0.237*	−0.158
	(0.133)	(0.134)	(0.135)	(0.128)	(0.131)	(0.133)	(0.121)	(0.121)	(0.126)	(0.136)	(0.137)
$\max t$		−0.012	−0.016		−0.010	−0.013		−0.020	−0.025		
		(0.022)	(0.021)		(0.021)	(0.020)		(0.021)	(0.020)		
$\max T$		0.036*	0.036*		0.029	0.028		0.039**	0.039**		
		(0.020)	(0.019)		(0.018)	(0.018)		(0.018)	(0.017)		
N			0.237***			0.207***			0.236***		
			(0.059)			(0.057)			(0.057)		
$\sigma^2_M(\theta_M)$	0.858	0.811	0.797	0.871	0.820	0.810	0.826	0.782	0.772	1.528	1.077
$\sigma^2_F(\theta_F)$	0.153	0.117	0.113	0.162	0.124	0.119	–	–	–	0.162	0.277
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (primary)							✓	✓	✓		
Year											✓

Note. Figures correspond to coefficients from fixed effects estimation of Equation (2). The dependent variable is citation counts (asinh). Both panels include co-authored papers, only. Panel one is estimated on the sample of male authors only; female co-author(s) is a binary variable equal to one if at least one co-author is female. Panel two is estimated on the sample of female authors only; male co-author(s) is a binary variable equal to one if at least one co-author is male. Standard errors clustered at the author level in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table 3: Returns to senior men from co-authoring with junior women

	(1)	(2)	(3)
female co-author	0.266	0.610***	0.603***
	(0.227)	(0.226)	(0.210)
$\max t$		−0.174***	−0.175***
		(0.029)	(0.027)
$\sigma^2_M(\theta_M)$	0.312	0.312	0.276
$\sigma^2_F(\theta_F)$	0.163	0.119	0.109
p-value (ratio)	0.000	0.000	0.000
No. obs.	314	314	314
Year×Journal	✓	✓	✓
Author	✓	✓	✓
JEL (primary)			✓

Note. Figures correspond to coefficients from fixed effects estimation of Equation (2) on the sub-sample of senior male authors with at least two top-five papers co-authored with exactly one economist of each sex who has no previous top-five publications. Female co-author is a binary variable equal to one if the junior co-author was female and 0 if he was male. Standard errors clustered at the author level in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

co-authors—*e.g.*, male economists may be more likely to collaborate with high-quality men on projects with at least one female co-author. To investigate, we limit our sample to papers and authors that satisfy the following criteria: senior male economists with at least two top-five papers published on or after 1990 that were co-authored with exactly one junior economist of each sex, where junior is defined has having no previous top-five publications.[14] The subsequent sub-sample yields one treatment group—56 senior men co-authoring with exactly one junior woman—and one control group—those same senior men co-authoring with exactly one junior man.[15]

The results, shown in Table 3, suggest that senior men's papers are higher quality when they are co-authored with junior women as opposed to junior men. In column (1), β_1 is comparable to estimates in Table 2, but its standard error is noticeably larger. After conditioning on $\max t$ and primary *JEL* fixed effects, however, it almost triples and becomes significant. ($\max T$ is perfectly collinear with senior author fixed effects and is therefore omitted as a control.) Thus, the results in Table 3 suggest that a senior man's paper is cited noticeably more if it is co-authored with a junior woman instead of a junior man.[16]

According to estimates in the second panel of Table 2, women's papers are *also* higher quality when they are co-authored with other women (Theorem 3.1, Condition 2). On average, women receive 13 log points fewer citations when they co-author with at least one man (column (1)). The gap falls an additional 10–18 log points after adjusting for $\max t$, $\max T$ and N. Results are similar when Equation (2) is estimated on the sample of women's papers published after 1990 and 2000 and controlling for *JEL*

[14] For example, Ariel Rubinstein co-authored "The 11–20 money request game: a level-k reasoning study" with Ayala Arad (*AER*, 2012) and "Back to fundamentals: equilibrium in abstract economics" with Michael Richter (*AER*, 2015). At the time of publication, Rubinstein had numerous previous top-five papers whereas Arad and Richter had none.

[15] Given the small number of senior authors, singleton groups are a particular problem when controlling for *JEL* fixed effects. In order to keep as many senior author groups in the estimation sample as possible, we duplicate articles by their number of *JEL* codes and assign each one a single code. Results and conclusions are similar—albeit less comparable across models—if we instead control for each *JEL* code separately as we do in Table 2 (see Hengel and Moon 2020, Table 3).

[16] This result contrasts with Card *et al.* (2020), who do not find a difference in citations accruing to mixed-gender papers with a senior male co-author compared to papers co-authored by all-male teams. As we show in Appendix D.9, we believe our conflicting results may be due to co-author composition effects that are less distortionary in the within-author comparisons shown in Table 3.

fixed effects.[17]

Results in Tables 2 and 3 suggest that variance in quality is lower when men *and* women co-author with women (Theorem 3.1, Condition 2). Estimates of the variance of ε_{it} in separate samples of papers by men that satisfy $\text{female}_{it} = 1$ and $\text{female}_{it} = 0$ persistently suggest that $\sigma^2_{iF}(\theta_{iF})$ is smaller than $\sigma^2_{iM}(\theta_{iM})$; estimates in papers by women satisfying $\text{male}_{it} = 1$ and $\text{male}_{it} = 0$ similarly indicate $\sigma^2_{iF}(\theta_{iF}) < \sigma^2_{iM}(\theta_{iM})$.[18]

As discussed in Section 4.2.2, evidence from other studies indicates that Condition 3 is likewise satisfied; we therefore tentatively conclude that $\theta_{iF} > \theta_{iM}$ for both male and female i. As emphasised in that section, however, we come to this conclusion only cautiously. First, Conditions 1–3 must actually hold *for the same* i; because we do not show this, the evidence we present in Tables 2 and 3 should be interpreted as suggestive, only. Second, in our opinion there are plausible scenarios in which Assumption 3 is violated, in which case our results are still informative about the *presence* of higher standards, but not about who, precisely, is responsible for setting them (for further discussion, see Section 4.2.2).

For robustness, we replicate Tables 2 and 3 using the log of 1 plus citations and raw counts as proxies for quality (Appendices D.1 and D.2). We also control non-parametrically for number of co-authors, institutional rank, secondary and tertiary *JEL* codes and account for the length of a manuscript's bibliography (Appendices D.3, D.4, D.5 and D.6). In all instances, the evidence supports our conclusion that female authors are likely held to higher standards compared to male authors.

6 Conclusion

Discrimination hurts its victims and, sometimes, its perpetrators (Becker 1957). For example, if an academic journal only publishes papers authored by men, its quality should decline relative to one that is gender-blind; if a male economist refuses to co-author with women, his papers ought to publish less well than the men who don't.

For these reasons, discrimination is sometimes considered incompatible with competitive forces. When markets are complete and perfect, the argument is roughly as follows: sufficient competition between unprejudiced journals should ensure female-authored papers are accepted at rates just equal to their marginal quality; sufficient competition between prejudiced and unprejudiced journals should ensure each is ranked according to the quality of the articles it publishes.[19] As long as a journal's ranking accurately prices its articles' quality, the quality of the papers it publishes should not vary by author gender nor should returns to co-authoring depend on a co-author's sex.[20]

According to our evidence, however, the articles top-five journals publish by women are higher quality than the articles they publish by men. Furthermore, both men and women publish higher quality papers when they co-author with women instead of men—for example, senior men receive almost 60 log points more citations per top-five paper when they co-author with junior women as opposed to junior men.

As we show in Theorem 3.1, higher quality female-authored papers (conditional on publication) could be

[17]The results in Tables 2 and 3 are much less sensitive to controlling for field than they were in Table 1. This may reveal an underlying association between field and author-specific unobservables that could partially bias estimates of β_1 downward when controlling for the former but not the latter in columns (7)–(9) of Table 1.

[18]The number of female authors with two or more exclusively female co-authored papers is too small to reliably estimate $\sigma^2_{iF}(\theta_{iF})$ when conditioning on *JEL* code; we therefore omit these results from Table 2.

[19]When markets are incomplete, however, even taste-based discrimination can persist in competitive equilibria or generate equilibria in which non-realised discrimination nevertheless results in an inefficient allocation of resources—*e.g.*, because some women flee to less discriminatory fields although their interests and talents would have been better matched to a career in economics. See, *e.g.*, Diamond (1971), Borjas and Bronars (1989) and Black (1995).

[20]If journal rankings perfectly price article quality, then every article published in the same journal must be exactly the same quality. If markets are complete but journal rankings *do not* precisely price article quality, then there must exist some other mechanism that does.

consistent with gender-neutral acceptance standards if women's papers are accepted more often or the variance in their quality is greater. Neither appears to be the case. Variance in quality is persistently lower in female-authored papers; evidence from a set of journals that partially overlaps with our own suggests men's and women's manuscripts are accepted at roughly equivalent rates (Card *et al.* 2020). Although there are several reasons to be cautious when interpreting our results—particularly when studying co-authored submissions by a single individual—on the balance of probabilities, we believe they point toward higher standards for female (co-)authors.

Unfortunately, our data cannot precisely identify *why* female authors are held to higher standards. The gender gaps we observe are not directly related to institutional prestige (Appendices C.6 and D.4) nor do they appear to be affected by rough controls for field, particularly after accounting for author-specific heterogeneity (see also additional analyses in Appendices C.7, D.5, C.8 and D.6). Yet there are many remaining channels. For example, referees may have biased beliefs about the quality of female-authored research. Alternatively (or additionally), women's smaller networks may mean editors and referees are less familiar with their work and consequently more risk averse about accepting it.

Another possibility is that referees use a paper's correspondence with some unobserved category—*e.g.*, number of equations—to proxy for quality. If female economists are more likely to submit manuscripts that fall outside these category boundaries, then referees will find their papers harder to evaluate and may therefore unconsciously penalise them during peer review.[21] Alternatively, there may be important disciplinary differences between male- and female-authored research; as a result, female economists could be more often disadvantaged by the "boundary maintenance" activities of predominantly male incumbents who fear women's research supplanting their their own. (For further discussion and evidence, see Fini *et al.* (2022); see also the related "pollution theory" of discrimination from Goldin (2014).)[22]

Ideally, publishing in a biased journal would send a weaker signal about the quality of male-authored papers than it would about the quality of female-authored papers. In the real world, however, expectations are slow to adjust (see, *e.g.*, Heckman and Moktan 2019). As a result, higher standards in peer review create higher standards for tenure and promotion. They also incentivise both genders to inefficiently collaborate with men.

In economics, we tend to favour policies targeted at individual market imperfections. But when the space of information asymmetries and transaction costs is large and poorly understood, active policy solutions—including formal and informal quotas—may be sensible alternatives (Lundberg 1991; Lundberg and Startz 1983). Not only are they non-punitive and verifiable, but they may also create positive externalities that could not have been achieved using markets alone (see, *e.g.*, Besley *et al.* 2017; Niederle *et al.* 2013). For example, clearly signalling a determination to publish more female authors will likely decrease the relative price of co-authoring with women and encourage more fruitful collaborations.

But active policy interventions are only Pareto improving when based on an adequate understanding of the context. More research is certainly needed. We hope journals are challenged to address the tougher standards they likely impose on women, willing to support the access and research needed to better understand them and open to whatever policy options most effectively check them.

References

Ahmadpoor, M. and B. F. Jones (2019). "Decoding team and individual inpact in sicence and invention". *PNAS* 116 (28), pp. 13885–13890.

[21] This is conceptually similar to the statistical discrimination modelled by Aigner and Cain (1977).

[22] "Boundary maintenance" may explain why the gender gap conditional on authors-specific fixed effects does not change after controlling for rough measures of field. In this case, referees would be effectively looking for a "cultural fit", which is better captured by individual fixed-effects and so no longer correlates with field after conditioning on them.

Aigner, D. J. and G. G. Cain (1977). “Statistical theories of discrimination in labor markets”. *ILR Review* 30 (2), pp. 175–187.

Aksnes, D. W. (2003). “Characteristics of highly cited papers”. *Research Evaluation* 12 (3), pp. 159–170.

Aksnes, D. W., L. Langfeldg, and P. Wouters (2019). “Citations, citation indicators and research quality: an overview of basic concepts and theories”. *SAGE Open* 9 (1), pp. 1–17.

Aksnes, D. W. and R. E. Taxt (2004). “Peer reviews and bibliometric indicators: a comparative study at a Norweigian university”. *Research Evaluation* 13 (1), pp. 33–41.

Alexander, D., O. Gorelkina, and E. Hengel (2021). “Gender and the time cost of peer review”. Mimeo.

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). “An evaluation of instrumental variable strategies for estimating the effects of Catholic schooling”. *Journal of Human Resources* 40 (4), pp. 791–821.

Anwar, S. and H. Fang (2006). “An alternative test of racial prejudice in motor vehicle searches: theory and evidence”. *American Economic Review* 96 (1), pp. 127–151.

Arnold, D., W. Dobbie, and C. S. Yang (2018). “Racial bias in bail decisions”. *Quarterly Journal of Economics* 133 (4), pp. 1885–1932.

Auriol, E. *et al.* (2022). “Underrepresentation of women in the economics profession more pronounced in the United States compared to heterogeneous Europe”. *PNAS* 119 (16), e2118853119.

Ayres, I. and J. Waldfogel (2006). “A market test for race discrimination in bail setting”. *Stanford Law Review* 46 (5), pp. 987–1047.

Azoulay, P., T. Stuart, and Y. Wang (2014). “Matthew: effect or fable?” *Management Science* 60 (1), pp. 92–109.

Ball, P. *et al.* (2020). “Gender issues in fundamental physics: Strumia’s bibliometric analysis fails to account for key confounders and confuses correlation with causation”. *Quantitative Science Studies* 2 (1), pp. 263–272.

Bateman, V., H. Gamage Danula Kankanam, and X. Liu (2021). *Royal Economic Society, Silver Anniversary Women’s Committee Report: The Gender Imbalance in UK Economics*. Report to the Royal Economic Society Women’s Committee.

Bateman, V. and E. Hengel (2022). “The gender gap in UK academic economics 1996–2018: progress, stagnation and retreat”. Mimeo.

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), pp. 221–242.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

– (1993). “The economic way of looking at behavior”. *Journal of Political Economy* 101 (3), pp. 385–409.

Besley, T. *et al.* (2017). “Gender quotas and the crisis of the mediocre man: theory and evidence from Sweden”. *American Economic Review* 107 (8), pp. 2204–2242.

Bjarnason, T. and I. D. Sigfusdottir (2002). “Nordic impact: article productivity and citation patterns in sixteen Nordic sociology departments”. *Acta Sociologica* 45 (4), pp. 253–267.

Black, D. A. (1995). “Discrimination in an equilibrium search model”. *Journal of Labor Economics* 13 (2), pp. 309–334.

Blank, R. M. (1991). “The effects of double-blind versus single-blind reviewing: experimental evidence from the *American Economic Review*”. *American Economic Review* 81 (5), pp. 1041–1067.

Borjas, G. J. and S. G. Bronars (1989). “Consumer discrimination and self-employment”. *Journal of Political Economy* 97 (3), pp. 581–605.

Card, D. *et al.* (2018). “Are referees and editors in economics gender neutral?” Mimeo.

– (2020). “Are referees and editors in economics gender neutral?” *Quarterly Journal of Economics* 135 (1), pp. 269–327.

Chari, A. and P. Goldsmith-Pinkham (2017). “Gender representation in economics across topics and time: evidence from the NBER Summer Institute”. NBER Working Paper No. 23953.

Chevalier, J. (2021). "Report from the Committee on the Status of Women in the Economics Profession (CSWEP)". *AEA Papers and Proceedings* 111, pp. 742–763.

D'Ippoliti, C. (2021). ""Many-citedness": citations measure more than just scientific quality". *Journal of Economic Surveys* 35 (5), pp. 1271–1301.

D'Ippoliti, C. *et al.* (2021). "There is strength in numbers. Scientific and professional ties as determinants of citations among UK economists". Rebuilding Macroeconomics Working Paper No. 60.

Diamond, P. A. (1971). "A model of price adjustment". *Journal of Economic Theory* 3 (2), pp. 156–168.

Dion, M. L., J. L. Sumner, and S. M. Mitchell (2018). "Gendered citation patterns across political science and social science methodology fields". *Political Analysis* 26 (3), pp. 312–327.

Ductor, L., S. Goyal, and A. Prummer (2021). "Gender and collaboration". *Review of Economics and Statistics* (forthcoming).

Dworkin, J. D. *et al.* (2020). "The extent and drivers of gender imbalance in neuroscience reference lists". *Nature Neuroscience* 23 (8), pp. 918–926.

Ferber, M. A. (1986). "Citations: are they an objective measure of scholarly merit?" *Signs* 11 (2), pp. 381–389.

– (1988). "Citations and networking". *Gender and Society* 2 (1), pp. 82–89.

Fini, R. *et al.* (2022). "A new take on the categorical imperative: gatekeeping, boundary maitenance, and evaluation penalities in sciene". *Organization Science* (forthcoming).

Foschi, M. (1996). "Double standards in the evaluation of men and women". *Social Psychology Quarterly* 59 (3), pp. 237–254.

Gamage, D. K., A. Sevilla, and S. Smith (2020). "Women in economics: a UK perspective". *Oxford Review of Economic Policy* 36 (4), pp. 962–982.

Gertsberg, M. (2022). "The unintended consequences of #MeToo: evidence from research collaborations". Mimeo.

Gibson, J., D. L. Anderson, and J. Tressler (2014). "Which journal rankings best explain academic salaries? Evidence from the University of California". *Economic Inquiry* 52 (4), pp. 1322–1340.

Ginther, D. K. and S. Kahn (2004). "Women in economics: moving up or falling off the academic career ladder?" *Journal of Economic Perspectives* 18 (3), pp. 193–214.

Goldin, C. (2014). "A pollution theory of discrimination: male and female differences in occupations and earnings". In: *Human Capital in History: The American Record*. University of Chicago Press, pp. 313–348.

Heckman, J. J. and S. Moktan (2019). "Publishing and promotion in economics: the tyranny of the Top Five". *Journal of Economic Literature* 38 (2), pp. 419–470.

Hengel, E. (2022). "Publishing while female. Are women held to higher standards? Evidence from peer review." *The Economic Journal* 132 (648), pp. 2951–2991.

Hengel, E. and E. Moon (2020). "Gender and quality at top economics journals". Mimeo.

Hospido, L. and C. Sanz (2021). "Gender gaps in the evaluation of research: evidence from submissions to economics conferences".

Knobloch-Westerwick, S., C. J. Glynn, and M. Huge (2013). "The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest". *Science Communication* 35 (5), pp. 603–625.

Knowles, J., N. Persico, and P. Todd (2001). "Racial Bias in Motor Vehicle Searches: Theory and Evidence". *Journal of Political Economy* 109 (1), pp. 203–229.

Koffi, M. (2019). "Innovative ideas and gender inequality". Mimeo.

– (2021). "Gendered citations at top economics journals". *AEA Papers and Proceedings* 111, pp. 60–64.

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, pp. 326–335.

Lundberg, S. J. (1991). "The enforcement of equal opportunity laws under imperfect information: affirmative action and alternatives". *Quarterly Journal of Economics* 106 (1), pp. 309–326.

Lundberg, S. J. and R. Startz (1983). "Private discrimination and social intervention in competitive labor markets". *American Economic Review* 73 (3), pp. 340–347.

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

Marx, P. (2021). "An absolute test of racial prejudice". *Journal of Law, Economics, and Organization* 38 (1), pp. 42–91.

Merton, R. K. (1968). "The Matthew effect in science". *Science* 159 (3810), pp. 56–63.

Moss-Racusin, C. A. *et al.* (2012). "Science faculty's subtle gender biases favor male students". *PNAS* 109 (41), pp. 16474–16479.

Niederle, M., C. Segal, and L. Vesterlund (2013). "How costly is diversity? Affirmative action in light of gender differences in competitiveness". *Management Science* 59 (1), pp. 1–16.

Oppenheim, C. (1997). "The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology". *Journal of Documentation* 53 (5), pp. 477–487.

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

Reuben, E., P. Sapienza, and L. Zingales (2014). "How stereotypes impair women's careers in science". *PNAS* 111 (12), pp. 4403–4408.

Rinia, E. J. *et al.* (1998). "Comparative analysis of a set of bibliometric indicators and central peer review criteria: evaluation of condensed matter physics in the Netherlands". *Research Policy* 27 (1), pp. 95–107.

Sarsons, H. *et al.* (2019). "Gender differences in recognition for group work". *Journal of Political Economy* 129 (1), pp. 101–147.

Simoiu, C., S. Corbett-Davies, and S. Goel (2017). "The problem of infra-marginality in outcome tests for discrimination". *Annals of Applied Statistics* 11 (3), pp. 1193–1216.

Teele, D. L. and K. Thelen (2017). "Gender in the journals: publication patterns in political science". *PS: Political Science and Politics* 50 (2), pp. 433–447.

van Raan, A. F. J. (2006). "Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups". *Scientometrics* 67 (3), pp. 491–502.

Waltman, L. (2016). "A review of the literature on citation impact indicators". *Journalof Informetrics* 10, pp. 365–391.

Zacchia, G. (2021). "What does it take to be top women economists? An analysis using rankings in RePEc". *Review of Political Economy* 33 (2), pp. 170–193.

Appendices

A Proofs

Proof of Theorem 3.1. Conditional on acceptance, the mean quality of papers by group $g \in \{M, F\}$ is

$$\mathbb{E}_g[q|q \geq \theta_g] = \int_{\theta_g}^{\infty} \frac{q\,\Phi_g'(q)}{1-\Phi_g(\theta_g)}\,\mathrm{d}q = \int_{\theta_g}^{\infty} \frac{1-\Phi_g(q)}{1-\Phi_g(\theta_g)}\,\mathrm{d}q + \theta_g, \tag{1}$$

where the last equality is obtained using integration by parts (see for example Hajeck (2015), p. 19; the Remark following this proof provides a full derivation). Thus,

$$\mathbb{E}_F[q|q \geq \theta_F] > \mathbb{E}_M[q|q \geq \theta_M]$$

is equivalent to

$$\int_{\theta_M}^{\infty} \frac{1-\Phi_M(q)}{1-\Phi_M(\theta_M)}\,\mathrm{d}q < \int_{\theta_M}^{\infty} \frac{1-\Phi_F(q)}{1-\Phi_F(\theta_F)}\,\mathrm{d}q - \int_{\theta_F}^{\theta_M} \frac{\Phi_F(q)-\Phi_F(\theta_F)}{1-\Phi_F(\theta_F)}\,\mathrm{d}q. \tag{2}$$

By way of a contradiction, assume $\theta_F \leq \theta_M$. Thus, $\Phi_F(\theta_F) \leq \Phi_F(q)$ for all $q \in (\theta_F, \theta_M)$, so Equation (2) together with $\Phi_M(\theta_M) = \Phi_F(\theta_F)$ implies

$$\int_{\theta_M}^{\infty} (1-\Phi_M(q))\,\mathrm{d}q < \int_{\theta_M}^{\infty} (1-\Phi_F(q))\,\mathrm{d}q. \tag{3}$$

Note that

$$\lim_{x\to\infty} \int_{y}^{x} \Phi_g(q)\,\mathrm{d}q = \infty \quad \text{for any } y \in \mathbb{R}. \tag{4}$$

Since Φ_F and Φ_M are continuous distributions, however, there exists a sufficiently large $\bar{q}$ such that Equation (3) implies

$$\int_{\theta_M}^{\bar{q}} \Phi_F(q)\,\mathrm{d}q < \int_{\theta_M}^{\bar{q}} \Phi_M(q)\,\mathrm{d}q. \tag{5}$$

Suppose $\sigma_M^2 = \sigma_F^2$. If $\mu_F \leq \mu_M$, then $\Phi_M(q) \leq \Phi_F(q)$ for all $q \in \mathbb{R}$, contradicting the inequality in Equation (5). But if $\mu_M < \mu_F$, $\Phi_F(q) < \Phi_M(q)$ for all $q \in \mathbb{R}$; combined with $\theta_F \leq \theta_M$, this implies

$$\Phi_F(\theta_F) \leq \Phi_F(\theta_M) < \Phi_M(\theta_M),$$

contradicting our assumption that $\Phi_F(\theta_F) = \Phi_M(\theta_M)$. Thus, $\sigma_M^2 \neq \sigma_F^2$.

Normal distributions are ordered in dispersion according to their variances (Lewis and Thompson 1981, Section 6.3). That is, the distribution with the greater variance dominates the other in the dispersive order (denoted by $<_{\text{disp}}$). $\Phi_g <_{\text{disp}} \Phi_{g'}$ and $\sigma_g^2 \neq \sigma_{g'}^2$ imply Φ_g intersects $\Phi_{g'}$ exactly once and from below (Shaked 1982, Theorem 2.1). Thus, $\Phi_{g'}(q) \leq \Phi_g(q)$ for all $q \geq q^\star$ where $q^\star < \infty$ uniquely satisfies $\Phi_g(q^\star) = \Phi_{g'}(q^\star)$.

If $q^\star \leq \theta_M$, then Equation (5) implies that Φ_M lies above Φ_F for all $q \geq q^\star$. To see that the same is true when $\theta_M < q^\star$, rewrite Equation (5) as

$$\int_{q^\star}^{\bar{q}} \Phi_F(q)\,\mathrm{d}q + \int_{\theta_M}^{q^\star} \Phi_F(q)\,\mathrm{d}q < \int_{q^\star}^{\bar{q}} \Phi_M(q)\,\mathrm{d}q + \int_{\theta_M}^{q^\star} \Phi_M(q)\,\mathrm{d}q. \tag{6}$$

As $\bar{q} \to \infty$, the limits of the first terms on each side of the inequality in Equation (6) are infinite (Equation

(4)) whereas the second terms are not. Thus, for a sufficiently large $\bar{q}'$, Equation (6) implies

$$\int_{q^\star}^{\bar{q}'} \Phi_F(q)\,\mathrm{d}q < \int_{q^\star}^{\bar{q}'} \Phi_M(q)\,\mathrm{d}q.$$

We therefore conclude that Φ_M lies above Φ_F for all $q \geq q^\star$. Thus, $\Phi_M <_{\text{disp}} \Phi_F$ and so $\sigma_M^2 < \sigma_F^2$ and also $\sigma_M^2(\theta_M) < \sigma_F^2(\theta_F)$ (without proof). This establishes the desired contradiction. □

Remark (Derivation of Equation 1). Recall from the first part of Equation (1) that

$$\begin{aligned}\mathbb{E}_g[q|q \geq \theta_g] &= \int_{\theta_g}^{\infty} \frac{q\,\Phi_g'(q)}{1-\Phi_g(\theta_g)}\,\mathrm{d}q \\ &= -\frac{1}{1-\Phi_g(\theta_g)} \int_{\theta_g}^{\infty} q\,\mathrm{d}(1-\Phi_g(q)).\end{aligned} \tag{7}$$

Using integration by parts on the last step of Equation (7), we get

$$\begin{aligned}&= -\frac{1}{1-\Phi_g(\theta_g)}\left(\lim_{q\to\infty}\{q(1-\Phi_g(q))\} - \theta_g\,(1-\Phi(\theta_g)) - \int_{\theta_g}^{\infty}(1-\Phi_g(q))\,\mathrm{d}q\right) \\ &= \int_{\theta_g}^{\infty} \frac{1-\Phi_g(q)}{1-\Phi_g(\theta_g)}\,\mathrm{d}q + \theta_g - \frac{1}{1-\Phi_g(\theta_g)} \lim_{q\to\infty} q(1-\Phi_g(q)).\end{aligned} \tag{8}$$

It remains to show that the limit in Equation (8) is zero. Note that

$$\lim_{q\to\infty} q(1-\Phi_g(q)) = \lim_{q\to\infty} \frac{1-\Phi_g(q)}{1/q}.$$

Applying l'Hôpital's rule, we have

$$\lim_{q\to\infty} \frac{1-\Phi_g(q)}{1/q} = \lim_{q\to\infty} \frac{\Phi_g'(q)}{1/q^2}. \tag{9}$$

Since Φ_g' is the density function for the normal distribution, Equation (9) is equivalent to

$$\begin{aligned}\lim_{q\to\infty} \frac{\Phi_g'(q)}{1/q^2} &= \lim_{q\to\infty} \frac{\frac{1}{\sqrt{2\pi\sigma_g^2}} \exp\left\{-\frac{(q-\mu_g)^2}{2\pi\sigma_g^2}\right\}}{1/q^2} \\ &= \frac{1}{\sqrt{2\pi\sigma_g^2}} \lim_{q\to\infty} \frac{q^2}{\exp\left\{\frac{(q-\mu_g)^2}{2\pi\sigma_g^2}\right\}} \\ &= 0.\end{aligned}$$

B Data

Data coverage. Our data include 10,951 full-length, original research articles published between 1950–2015 in the *AER*, *ECA*, *JPE*, *QJE* and *REStud*.[1] We define "full-length, original research" as any article published with an abstract, excluding articles published in the *Papers & Proceedings* issues of the *AER*, errata and corrigenda. We make this distinction because before 1990, almost all top-five journals—and especially *JPE* and *AER*—published a large variety of non-original research—*e.g.*, book reviews, editorials and reports—that rarely included an abstract.

Data coverage by journal and decade are shown in Table B.1. Before 1980, our dataset includes only articles published in *ECA*, *JPE* and *REStud*—these journals systematically published abstracts with their full-length, original research articles before the *AER* and *QJE*. Starting in the mid-1980s, however, almost all original research published in any top-five journal contained an abstract and are therefore included in our data.

Table B.1: Data coverage by journal and decade.

Decade	*AER*	*ECA*	*JPE*	*QJE*	*REStud*	Total
1950-59		120				120
1960-69		344	184			528
1970-79		660	633	1	227	1,521
1980-89	180	648	562	401	490	2,281
1990-99	476	443	478	409	383	2,189
2000-09	693	519	408	413	430	2,463
2010-15	732	382	181	251	303	1,849
Total	2,081	3,116	2,446	1,475	1,833	10,951

Author gender. Each of the 7,559 unique authors in our dataset was manually assigned a gender based on (i) obviously gendered given names (*e.g.*, "James" or "Brenda"); (ii) photographs on personal or faculty websites; (iii) personal pronouns used in text written about the individual; and (iv) by contacting the author himself or people and institutions connected to him.

Citation source data. Citation data were obtained from Web of Science (2018), a comprehensive database of all social science research published since 1900. Counts correspond to the number of published papers in the Web of Science database that cite a given article and include self-citations to later work. Citations for *AER*, *ECA*, *JPE* and *QJE* were first collected in August 2017 and updated in January 2018; citations for *REStud* were collected in October 2018.

Independent variable definitions. Table B.2 specifies precisely how each of the independent variables used in the analysis were calculated.

[1]The data were originally collected and analysed in Hengel (2022) and Hengel (2017). The original dataset analysed in Hengel (2017) included only articles published with an abstract between 1950–2015 in the *AER*, *ECA*, *JPE* and *QJE*. Later, Hengel (2022) added full-length articles published with a submit-accept date in *REStud*. (Almost all of these articles also include an abstract, but the presence of a submit-accept date is effectively another indicator that an article is original research and fully peer reviewed.)

Table B.2: Variable descriptions

Variable name	Description	Notes
N	Number of co-authors on a paper.	
$\max t$	Author seniority at the time of publication, measured as the number of top-five papers the most prolific author on a paper had at the time the paper was published.	For example, consider the paper "Efficient intra-household allocations: a general characterization and empirical tests" by Martin J. Browning and Pierre-André Chiappori, which was published in *Econometrica* in 1998. In 1998, Chiappori had 7 top-five publications (including this paper) and Browning had 11, so $\max t = \max(7, 11) = 11$. By the end of 2015, however, Browning had 19 top-five publications whereas Chiappori had 21. Thus, $\max T = \max(19, 21) = 21$.
$\max T$	Author prominence at the time citations were collected, measured as the total number of top-five papers the most prolific author on a paper had at the end of 2015.	
JEL fixed effects	Fixed effects for *JEL* codes. In the body of the paper, we control for primary *JEL* code fixed effects; in Appendices C.7 and D.5 we control for secondary and tertiary *JEL* code fixed effects.	*JEL* codes were significantly revised in 1990; comparable codes are not available for periods pre- and post-reform. We therefore only control for *JEL* fixed effects in articles published after the reform.
Institutional rank fixed effects	Fixed effects for the rank of authors' institutions.	To determine institutional rank, we follow the same procedure as Hengel (2022). That is, for each institution, we count the number of articles in which it was listed as an affiliation in a given year and smooth the average over a five-year period. Institutions are ranked on an annual basis using this figure and then grouped to create fifteen dynamic dummy variables. Institutions ranked in positions 1–9 are assigned individual dummy variables. Those in positions 10–59 are grouped in bins of 10 to form six dummy variables. Institutions ranked 60 or above were collectively grouped to form a final dummy variable. When multiple institutions are associated with an article, only the dummy variable of the highest ranked institution is used. For more details on how data for institutions were collected, see Hengel (2022).
Bibliography length	Number of works cited in a manuscript's bibliography.	Data from Web of Science.

C Section 5.1, supplemental output

C.1 Quantile regression

Table C.1 re-estimates Table 1 using a quantile regression model and raw citation counts as the dependent variable. The first panel replicates Table 1, column (3) at the 25th, median and 75th percentiles of citations; the second panel similarly replicates column (9). The coefficient on female is positive across all three percentiles, but standard errors are larger in the 75th percentile.

Table C.1: Table 1 columns (6) and (9), quantile regression

	without *JEL* fixed effects			with *JEL* fixed effects		
	25 pc.	50 pc.	75 pc.	25 pc.	50 pc.	75 pc.
	(1)	(2)	(3)	(4)	(5)	(6)
female	3.000***	5.666***	4.400	3.463***	4.745***	1.802
	(0.915)	(1.396)	(2.705)	(1.094)	(1.755)	(2.932)
N	2.767***	4.589***	5.800***	2.901***	4.520***	6.144***
	(0.4)	(0.667)	(1.092)	(0.523)	(0.792)	(1.401)
$\max t$	−1.438***	−3.002***	−6.000***	−2.779***	−4.290***	−7.328***
	(0.164)	(0.227)	(0.668)	(0.231)	(0.432)	(0.941)
$\max T$	1.444***	3.079***	6.600***	2.716***	4.413***	7.617***
	(0.138)	(0.196)	(0.582)	(0.203)	(0.41)	(0.846)
Constant	4.919	5.940	58.600	10.284	60.726***	151.848***
	(5.688)	(20.669)	(46.206)	(9.593)	(22.511)	(40.675)
$\sigma^2_M(\theta_M)$	117,741	115,873	113,848	51,140	49,097	47,745
$\sigma^2_F(\theta_F)$	18,424	16,713	19,688	19,943	17,930	20,156
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921
Year×Journal	✓	✓	✓	✓	✓	✓
JEL (primary)				✓	✓	✓

Note. First panel replicates results shown in Table 1, column (3) across different percentiles of the distribution using quantile regressions and raw citation counts as the dependent variable; second panel similarly replicates results from column (9). Bootstrapped standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.2 Negative binomial

In Table C.2, we estimate Equation (1) in a negative binomial model. Similar to results in Table 1, β_1 is small and insignificant before controlling for the Matthew effect. After controlling for $\max t$ and $\max T$, however, it becomes positive and significant.

In the final panel of Table C.2, we restrict the sample to include only papers published between 2000–2015 and omit controls for N, $\max t$ and $\max T$. Again, results suggest that women's papers are, on average, higher quality conditional on publication.

Theorem 3.1 does not apply if submissions are assumed to follow a negative binomial distribution. We therefore do not estimate the variance of male- and female-authored paper quality, conditional on acceptance.

Table C.2: Table 1, negative binomial model

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female	−0.002	0.008	0.086**	0.037	0.054	0.109***	−0.005	0.009	0.067*	0.105***	0.155***
	(0.04)	(0.04)	(0.039)	(0.038)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.04)	(0.041)
N		0.179***	0.139***		0.187***	0.158***		0.184***	0.151***		
		(0.016)	(0.017)		(0.017)	(0.017)		(0.016)	(0.017)		
$\max t$			−0.050***			−0.036***			−0.036***		
			(0.003)			(0.004)			(0.004)		
$\max T$			0.054***			0.040***			0.042***		
			(0.002)			(0.003)			(0.003)		
Constant	2.744***	2.392***	2.416***	2.735***	2.366***	2.370***	2.585***	2.235***	2.265***	2.721***	5.019***
	(0.185)	(0.187)	(0.184)	(0.155)	(0.158)	(0.156)	(0.16)	(0.162)	(0.16)	(0.144)	(0.066)
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (primary)							✓	✓	✓		
Year											✓

Note. Figures correspond to coefficients from estimating Equation (1) in a negative binomial model with raw citation counts as the dependent variable. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.3 Log of 1 plus citations

Table C.3 replicates Table 1 using the log of 1 plus citations as the dependent variable. As expected, the results are very similar to the results shown in Table 1.

Table C.3: Table 1, log of 1 plus citations

	All data			1990–2015						2000–2015	
				without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female (β_1)	0.070*	0.083**	0.154***	0.085**	0.106***	0.163***	0.027	0.047	0.109***	0.131***	0.173***
	(0.036)	(0.036)	(0.035)	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)	(0.036)	(0.041)	(0.042)
N		0.206***	0.163***		0.190***	0.159***		0.186***	0.154***		
		(0.016)	(0.016)		(0.017)	(0.017)		(0.017)	(0.017)		
$\max t$			−0.046***			−0.043***			−0.042***		
			(0.004)			(0.004)			(0.004)		
$\max T$			0.051***			0.047***			0.047***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.548	1.528	1.444	1.109	1.088	1.037	1.058	1.039	0.988	0.922	0.998
$\sigma^2_F(\theta_F)$	0.821	0.805	0.779	0.797	0.779	0.754	0.764	0.747	0.721	0.747	0.897
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.088
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
R^2	0.264	0.274	0.313	0.348	0.361	0.390	0.375	0.387	0.416	0.402	0.353
Bounds (β_1)	[0.03,0.07]	[0.06,0.08]	[0.15,0.20]	[0.09,0.21]	[0.11,0.25]	[0.16,0.37]	[0.03,0.10]	[0.05,0.14]	[0.11,0.27]	[0.13,0.18]	[0.17,0.27]
Year×Jnl	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1 except that the dependent variable is the log of 1 plus citations. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.4 Alternative proxies for article gender

The following tables replicate Table 1 using alternative definitions of female authorship. Table C.4 compares papers with a senior female author to papers with a senior male author. Table C.5 replaces a binary variable of female authorship with a continuous measure of the ratio of female authors on a paper. In Table C.6, we define female-authorship as in Table 1, but also include mixed gendered papers with fewer than 50 percent female co-authors and classify them as male-authored papers. In Table C.7 we compare papers with at least one female author to papers that are exclusively male-authored. Table C.8 restricts the sample to solo-authored papers, only. Finally, Table C.9 compares entirely male co-authored papers to entirely female co-authored papers. Mixed-gendered papers not satisfying the relevant "female" criteria in Table C.4, all co-authored papers in Table C.8, and all solo-authored and mixed-gendered co-authored papers in Table C.9 are dropped.

In general, results in Tables C.4–C.9 are similar to those presented in Table 1, especially after accounting for the Matthew effect.

See Hengel and Moon (2020, Table D.5) for similar results using a categorical variable to account for seven different gender categories: (i) female solo-authored, (ii) female co-authored, (iii) mixed sex co-authored with a senior female author, (iv) mixed sex co-authored with senior male and female authors of equal rank, (v) mixed sex co-authored with a senior male author, (iv) male solo-authored and (vii) male co-authored.

Table C.4: Table 1, senior female author

	1990–2015										
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
sr. fem. (β_1)	0.001	0.033	0.134***	0.028	0.062	0.145***	−0.035	−0.003	0.087**	0.080*	0.137***
	(0.042)	(0.042)	(0.042)	(0.043)	(0.043)	(0.042)	(0.042)	(0.042)	(0.042)	(0.047)	(0.048)
N		0.219***	0.176***		0.202***	0.170***		0.193***	0.161***		
		(0.015)	(0.015)		(0.016)	(0.016)		(0.016)	(0.016)		
$\max t$			−0.047***			−0.043***			−0.041***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.715	1.688	1.600	1.211	1.183	1.130	1.157	1.132	1.079	1.033	1.118
$\sigma^2_F(\theta_F)$	0.783	0.773	0.736	0.753	0.738	0.701	0.714	0.701	0.661	0.722	0.943
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016
No. obs.	10,951	10,951	10,951	6,285	6,285	6,285	6,285	6,285	6,285	4,312	4,312
R^2	0.269	0.280	0.318	0.353	0.367	0.396	0.380	0.394	0.422	0.406	0.357
Bounds (β_1)	[-0.04,0.00]	[0.02,0.03]	[0.13,0.23]	[0.03,0.15]	[0.06,0.22]	[0.15,0.40]	[-0.04,0.03]	[0.00,0.09]	[0.09,0.29]	[0.08,0.11]	[0.14,0.22]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced with a dummy variable equal to 1 if a paper had at least one female author who had previously published at least as many top-five papers as her co-authors at the time the paper paper in question was published. (Mixed-gendered papers with a senior male co-author are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.5: Table 1, ratio of female authors

	1990–2015										
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
fem. ratio (β_1)	0.092*	0.115**	0.227***	0.123**	0.150***	0.240***	0.028	0.055	0.151***	0.202***	0.267***
	(0.051)	(0.051)	(0.05)	(0.053)	(0.052)	(0.052)	(0.052)	(0.052)	(0.051)	(0.06)	(0.061)
N		0.219***	0.176***		0.202***	0.170***		0.194***	0.161***		
		(0.015)	(0.015)		(0.016)	(0.016)		(0.016)	(0.016)		
$\max t$			−0.048***			−0.044***			−0.042***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.747	1.725	1.633	1.216	1.193	1.139	1.160	1.139	1.085	1.034	1.117
$\sigma^2_F(\theta_F)$	0.653	0.648	0.631	0.631	0.621	0.608	0.525	0.521	0.513	0.610	0.952
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.141
No. obs.	10,951	10,951	10,951	6,285	6,285	6,285	6,285	6,285	6,285	4,312	4,312
R^2	0.269	0.280	0.318	0.353	0.368	0.397	0.380	0.394	0.422	0.407	0.359
Bounds (β_1)	[0.06,0.09]	[0.10,0.11]	[0.23,0.33]	[0.12,0.34]	[0.15,0.39]	[0.24,0.58]	[0.03,0.16]	[0.05,0.21]	[0.15,0.42]	[0.20,0.33]	[0.27,0.46]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced with a continuous variable equal to the ratio of female authors on a paper. ($\sigma^2_M(\theta_M)$ is estimated on the sample of exclusively male-authored papers and $\sigma^2_F(\theta_F)$ is estimated on the sample of exclusively female-authored papers.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.6: Table 1, 50 percent female authors

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
50% female (β_1)	0.057	0.090**	0.163***	0.068*	0.111***	0.169***	0.010	0.052	0.115***	0.110**	0.153***
	(0.037)	(0.037)	(0.037)	(0.039)	(0.039)	(0.038)	(0.039)	(0.039)	(0.038)	(0.043)	(0.044)
N		0.220***	0.178***		0.204***	0.173***		0.195***	0.163***		
		(0.015)	(0.015)		(0.016)	(0.016)		(0.016)	(0.016)		
$\max t$			−0.048***			−0.044***			−0.042***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.726	1.700	1.609	1.210	1.182	1.127	1.153	1.129	1.074	1.029	1.115
$\sigma^2_F(\theta_F)$	0.912	0.894	0.866	0.881	0.860	0.833	0.845	0.827	0.798	0.833	0.996
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.069
No. obs.	10,951	10,951	10,951	6,285	6,285	6,285	6,285	6,285	6,285	4,312	4,312
R^2	0.269	0.280	0.318	0.353	0.368	0.397	0.380	0.394	0.422	0.407	0.358
Bounds (β_1)	[0.00,0.06]	[0.06,0.09]	[0.16,0.21]	[0.07,0.17]	[0.11,0.25]	[0.17,0.37]	[0.01,0.05]	[0.05,0.14]	[0.12,0.27]	[0.11,0.14]	[0.15,0.23]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced with a dummy variable equal to 1 if at least 50 percent of authors are female and 0 otherwise. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.7: Table 1, at least one female author

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1+ female (β_1)	0.139***	0.074**	0.146***	0.158***	0.098***	0.152***	0.097***	0.038	0.095***	0.194***	0.234***
	(0.033)	(0.033)	(0.033)	(0.034)	(0.034)	(0.034)	(0.034)	(0.034)	(0.033)	(0.037)	(0.038)
N		0.214***	0.166***		0.194***	0.157***		0.191***	0.153***		
		(0.016)	(0.016)		(0.016)	(0.016)		(0.016)	(0.016)		
$\max t$			−0.048***			−0.044***			−0.042***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.747	1.725	1.633	1.216	1.193	1.139	1.160	1.139	1.085	1.034	1.117
$\sigma^2_F(\theta_F)$	0.965	0.941	0.900	0.931	0.907	0.868	0.892	0.870	0.833	0.861	1.002
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.039
No. obs.	10,951	10,951	10,951	6,285	6,285	6,285	6,285	6,285	6,285	4,312	4,312
R^2	0.270	0.280	0.318	0.355	0.368	0.397	0.381	0.394	0.422	0.409	0.361
Bounds (β_1)	[0.14,0.16]	[0.03,0.07]	[0.15,0.18]	[0.16,0.37]	[0.10,0.26]	[0.15,0.38]	[0.10,0.26]	[0.04,0.14]	[0.10,0.27]	[0.19,0.32]	[0.23,0.40]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced with a dummy variable equal to 1 if at least one author on a paper is female. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.8: Table 1, solo-authored papers

			1990–2015					
	All data		without *JEL* fixed effects		with *JEL* fixed effects		2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
solo female (β_1)	0.079	0.195***	0.108	0.189**	0.009	0.095	0.165*	0.151*
	(0.071)	(0.071)	(0.075)	(0.075)	(0.073)	(0.073)	(0.091)	(0.089)
$\max t$		−0.050***		−0.050***		−0.045***		
		(0.007)		(0.009)		(0.009)		
$\max T$		0.060***		0.056***		0.053***		
		(0.004)		(0.007)		(0.007)		
$\sigma^2_M(\theta_M)$	2.001	1.891	1.222	1.155	1.157	1.096	0.986	1.104
$\sigma^2_F(\theta_F)$	0.655	0.635	0.630	0.613	0.501	0.491	0.609	0.988
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.383
No. obs.	4,874	4,874	1,916	1,916	1,916	1,916	1,106	1,106
R^2	0.220	0.262	0.315	0.350	0.352	0.383	0.390	0.327
Bounds (β_1)	[-0.03,0.08]	[0.20,0.20]	[0.11,0.33]	[0.19,0.51]	[0.01,0.15]	[0.09,0.33]	[0.17,0.36]	[0.15,0.33]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)					✓	✓		
Year								✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced with a dummy variable equal to 1 if the paper was solo-authored by a woman and zero if it was solo-authored by a man. (Co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.9: Table 1, 100 percent female co-authored papers

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
100% fem. (β_1)	0.260**	0.311**	0.394***	0.275**	0.341***	0.419***	0.157	0.219*	0.302***	0.329***	0.495***
	(0.129)	(0.129)	(0.127)	(0.118)	(0.117)	(0.116)	(0.118)	(0.118)	(0.117)	(0.111)	(0.112)
N		0.170***	0.160***		0.213***	0.201***		0.202***	0.191***		
		(0.03)	(0.03)		(0.03)	(0.03)		(0.029)	(0.029)		
$\max t$			−0.039***			−0.037***			−0.037***		
			(0.005)			(0.006)			(0.006)		
$\max T$			0.043***			0.042***			0.042***		
			(0.003)			(0.004)			(0.004)		
$\sigma_M^2(\theta_M)$	1.382	1.374	1.310	1.140	1.125	1.081	1.085	1.072	1.027	1.004	1.099
$\sigma_F^2(\theta_F)$	0.158	0.152	0.115	0.168	0.162	0.123	0.000	0.000	0.000	0.153	0.224
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	5,048	5,048	5,048	3,460	3,460	3,460	3,460	3,460	3,460	2,444	2,444
R^2	0.344	0.348	0.378	0.403	0.411	0.434	0.432	0.439	0.462	0.434	0.380
Bounds (β_1)	[0.26,0.36]	[0.31,0.46]	[0.39,0.63]	[0.28,0.44]	[0.34,0.57]	[0.42,0.73]	[0.16,0.20]	[0.22,0.33]	[0.30,0.50]	[0.33,0.37]	[0.49,0.70]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that the independent variable female has been replaced by a dummy variable equal to 1 if the paper was co-authored entirely by women and 0 if it was co-authored entirely by men. (Solo-authored and mixed-gendered co-authored papers are excluded.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.5 Controlling non-parametrically for the number of co-authors

Table C.10 replicates Table 1 but controls non-parametrically for the number of co-authors. (Given space constraints, we do not report the coefficients on each fixed effect for number of co-authors.) Results are very similar to those reported in Table 1.

Table C.10: Table 1, controlling non-parametrically for the number of co-authors

				1990–2015							
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female (β_1)	0.077**	0.079**	0.160***	0.093**	0.112***	0.176***	0.031	0.048	0.117***	0.141***	0.184***
	(0.038)	(0.038)	(0.037)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.043)	(0.044)
$\max t$			−0.048***			−0.044***			−0.043***		
			(0.004)			(0.005)			(0.005)		
$\max T$			0.053***			0.049***			0.048***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.747	1.724	1.633	1.216	1.193	1.138	1.160	1.138	1.085	1.034	1.117
$\sigma^2_F(\theta_F)$	0.912	0.892	0.862	0.881	0.858	0.830	0.845	0.825	0.796	0.833	0.996
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.067
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
R^2	0.265	0.275	0.312	0.349	0.362	0.391	0.376	0.389	0.417	0.403	0.355
Bounds (β_1)	[0.03,0.08]	[0.03,0.08]	[0.16,0.20]	[0.09,0.22]	[0.11,0.26]	[0.18,0.40]	[0.03,0.10]	[0.05,0.14]	[0.12,0.29]	[0.14,0.20]	[0.18,0.28]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
N		✓	✓		✓	✓		✓	✓		
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1 except that models in (2), (3), (5), (6), (8) and (9) control non-parametrically for N. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.6 Controlling for institutional rank

Table C.11 replicates Table 1 but includes fixed effects for institutional rank. (See Appendix B for information on how institutional rank was constructed.) Results are very similar to those in Table 1.

Table C.11: Table 1, controlling for institutional rank

	1990–2015										
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female (β_1)	0.062*	0.073**	0.123***	0.072*	0.089**	0.131***	0.019	0.035	0.081**	0.124***	0.161***
	(0.037)	(0.037)	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.038)	(0.042)	(0.043)
N		0.159***	0.145***		0.149***	0.138***		0.148***	0.137***		
		(0.017)	(0.017)		(0.018)	(0.018)		(0.017)	(0.018)		
$\max t$			−0.046***			−0.041***			−0.039***		
			(0.004)			(0.004)			(0.004)		
$\max T$			0.043***			0.040***			0.040***		
			(0.003)			(0.003)			(0.003)		
$\sigma^2_M(\theta_M)$	1.633	1.622	1.571	1.132	1.121	1.090	1.087	1.076	1.045	0.972	1.039
$\sigma^2_F(\theta_F)$	0.855	0.842	0.827	0.822	0.805	0.791	0.790	0.775	0.761	0.764	0.900
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
R^2	0.312	0.318	0.338	0.394	0.401	0.417	0.415	0.422	0.438	0.439	0.401
Bounds (β_1)	[0.00,0.06]	[0.02,0.07]	[0.12,0.12]	[0.07,0.18]	[0.09,0.22]	[0.13,0.31]	[0.02,0.08]	[0.04,0.11]	[0.08,0.21]	[0.12,0.16]	[0.16,0.24]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Institution	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that all models include fixed effects for the institutional rank of the author from the highest ranked institution. (See Appendix B for information on how institutional rank was constructed.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.7 Controlling for secondary and tertiary *JEL* codes

Table C.12 replicates columns (7)–(9) in Table 1 but includes fixed effects for secondary (columns (1)–(3)) and tertiary (columns (4)–(6)) *JEL* codes. The coefficients on female are very similar to the estimates that control for primary *JEL* code fixed effects reported in Table 1.

Table C.12: Table 1, controlling for secondary and tertiary *JEL* codes

	Secondary *JEL* fixed effects			Tertiary *JEL* fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
female (β_1)	0.036	0.057	0.114***	0.021	0.045	0.104**
	(0.04)	(0.039)	(0.038)	(0.043)	(0.042)	(0.041)
N		0.198***	0.165***		0.199***	0.161***
		(0.018)	(0.018)		(0.019)	(0.019)
$\max t$			−0.041***			−0.041***
			(0.005)			(0.005)
$\max T$			0.046***			0.048***
			(0.003)			(0.004)
$\sigma^2_M(\theta_M)$	1.096	1.075	1.027	0.953	0.934	0.887
$\sigma^2_F(\theta_F)$	0.716	0.700	0.674	0.324	0.318	0.305
p-value	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	5,921	5,921	5,921	5,921	5,921	5,921
R^2	0.410	0.422	0.447	0.480	0.491	0.517
Bounds (β_1)	[0.04,0.12]	[0.06,0.16]	[0.11,0.29]	[0.02,0.11]	[0.05,0.18]	[0.10,0.31]
Year×Journal	✓	✓	✓	✓	✓	✓
JEL (secondary)	✓	✓	✓			
JEL (tertiary)				✓	✓	✓

Note. Estimates are identical to those in columns (7)–(9) in Table 1, except that columns (1)–(3) include fixed effects for secondary *JEL* categories and columns (4)–(6) include fixed effects for tertiary *JEL* categories. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.8 Controlling for length of the bibliography

Tables 1 and C.12 account for field using primary, secondary and tertiary *JEL* fixed effects. A limitation of this approach is that it crucially depends on the accuracy of the *JEL* classification system, which, unfortunately, "does not provide a pure image of the discipline" (Cherrier 2017, p. 547).[2]

An alternative approach common in the bibliometric literature is citing-side normalisation. Citing-side normalisation techniques aim to account for field-specific differences in the propensity to cite. To understand how this can distort estimates of gender differences in citations at the mean, suppose there are two fields: field A is female-dominated and field B is male-dominated. In both fields there are 100 researchers, each researcher has authored exactly one paper and the "quality" of every paper is exactly the same. However, the custom in field A is for every researcher to cite all 99 other papers in A, whereas the custom in field B is to randomly cite only 9 other papers in B. Thus, every paper in A receives 99 citations but papers in B are only cited (on average) by 9 other papers. Given A is female-dominated and B is male-dominated, estimates of gender differences in citations will give an inaccurate picture of the true gender difference in quality.

If A and B are clearly defined and observed by the researcher, then the obvious solution is simply to condition on them directly. In most situations, however, the boundaries between fields are poorly defined and difficult to observe. Citing-side normalisations circumvents this problem by accounting for the field-specific citation patterns themselves. This concept originates from Zitt and Small (2008) and numerous citing-side normalisation techniques have since emerged (for a discussion and references, see Waltman 2016).

In this section, we take a straightforward approach and simply control for the number of papers listed in each article's reference list. Results are shown in Table C.13. Consistent with other studies, papers with longer reference lists are also cited more, on average. However, controlling for bibliography length does not appear to affect the direction—and has only a small impact on the magnitude—of the coefficient on female.

[2] Cherrier (2017) provides a fascinating historical account of the evolution and limitations of and controversies surrounding the *JEL* classification system.

Table C.13: Table 1, controlling for length of the bibliography

	1990–2015									2000–2015	
	All data			without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
female (β_1)	0.063*	0.077**	0.149***	0.068*	0.089**	0.145***	0.020	0.040	0.100***	0.106**	0.139***
	(0.037)	(0.037)	(0.037)	(0.039)	(0.038)	(0.038)	(0.039)	(0.038)	(0.038)	(0.042)	(0.043)
N		0.212***	0.169***		0.191***	0.161***		0.183***	0.152***		
		(0.016)	(0.016)		(0.017)	(0.017)		(0.017)	(0.017)		
$\max t$			−0.046***			−0.039***			−0.038***		
			(0.004)			(0.004)			(0.005)		
$\max T$			0.051***			0.044***			0.043***		
			(0.003)			(0.003)			(0.003)		
bibl. length	0.022***	0.022***	0.021***	0.018***	0.018***	0.017***	0.019***	0.019***	0.018***	0.017***	0.019***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\sigma^2_M(\theta_M)$	1.636	1.615	1.533	1.113	1.093	1.048	1.059	1.041	0.997	0.929	0.986
$\sigma^2_F(\theta_F)$	0.887	0.868	0.840	0.854	0.832	0.805	0.822	0.803	0.774	0.801	0.948
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.535
No. obs.	10,566	10,566	10,566	5,921	5,921	5,921	5,921	5,921	5,921	3,982	3,982
R^2	0.309	0.318	0.352	0.399	0.411	0.434	0.425	0.436	0.459	0.456	0.423
Bounds (β_1)	[0.00,0.06]	[0.03,0.08]	[0.15,0.18]	[0.07,0.17]	[0.09,0.21]	[0.14,0.33]	[0.02,0.08]	[0.04,0.12]	[0.10,0.25]	[0.11,0.13]	[0.14,0.19]
Year×Jnl.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
JEL (prim.)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 1, except that all models control for the length of a paper's bibliography. (See Appendix B for information on how this indicator was constructed.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D Section 5.2, supplemental output

D.1 Log of 1 plus citations

Tables D.1 and D.2 replicate Tables 2 and 3, respectively, but use the log of 1 plus citations as the dependent variables. Again—and as expected—results are very similar to those presented in Tables 2 and 3.

Table D.1: Table 2, log of 1 plus citations

	All data			1990–2015 without *JEL* fixed effects			1990–2015 with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	0.104***	0.119***	0.076**	0.112***	0.124***	0.067*	0.084**	0.096***	0.041	0.136***	0.142***
	(0.036)	(0.036)	(0.036)	(0.037)	(0.037)	(0.037)	(0.036)	(0.036)	(0.037)	(0.038)	(0.038)
$\max t$		−0.011***	−0.012***		−0.014***	−0.016***		−0.014***	−0.016***		
		(0.004)	(0.004)		(0.004)	(0.004)		(0.004)	(0.004)		
$\max T$		0.020***	0.019***		0.023***	0.021***		0.023***	0.022***		
		(0.003)	(0.003)		(0.004)	(0.004)		(0.004)	(0.004)		
N			0.112***			0.136***			0.133***		
			(0.019)			(0.019)			(0.018)		
$\sigma^2_M(\theta_M)$	0.499	0.495	0.494	0.394	0.390	0.387	0.384	0.380	0.376	0.306	0.326
$\sigma^2_F(\theta_F)$	0.116	0.115	0.114	0.114	0.112	0.111	0.106	0.105	0.104	0.118	0.157
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−0.129	−0.227*	−0.304**	−0.189	−0.262**	−0.330**	−0.142	−0.222*	−0.317***	−0.233*	−0.150
	(0.129)	(0.129)	(0.130)	(0.124)	(0.127)	(0.128)	(0.116)	(0.116)	(0.121)	(0.130)	(0.133)
$\max t$		−0.014	−0.018		−0.011	−0.015		−0.022	−0.027		
		(0.021)	(0.020)		(0.019)	(0.019)		(0.020)	(0.019)		
$\max T$		0.036*	0.036**		0.029	0.028*		0.039**	0.039**		
		(0.019)	(0.018)		(0.018)	(0.017)		(0.017)	(0.016)		
N			0.231***			0.202***			0.229***		
			(0.057)			(0.054)			(0.055)		
$\sigma^2_M(\theta_M)$	0.774	0.731	0.718	0.784	0.738	0.728	0.744	0.703	0.693	1.366	0.975
$\sigma^2_F(\theta_F)$	0.147	0.112	0.108	0.156	0.119	0.114	–	–	–	0.156	0.264
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (primary)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 2 except that the dependent variable is the log of 1 plus citations. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.2: Table 3, log of 1 plus citations

	(1)	(2)	(3)
female co-author	0.228	0.575**	0.569***
	(0.223)	(0.222)	(0.207)
$\max t$		−0.176***	−0.177***
		(0.029)	(0.028)
$\sigma^2_M(\theta_M)$	0.290	0.290	0.258
$\sigma^2_F(\theta_F)$	0.147	0.106	0.097
p-value (ratio)	0.000	0.000	0.000
No. obs.	314	314	314
Year×Journal	✓	✓	✓
Author	✓	✓	✓
JEL (primary)			✓

Note. Estimates are identical to those in Table 3 except that the dependent variable is the log of 1 plus citations. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.2 Raw citation counts

Tables D.3 and D.4 replicate Tables 2 and 3, but use raw citation counts as the dependent variable. The coefficients on g_{it}^{-i} in Tables D.3 and D.4 are almost always in the same direction as the corresponding coefficients in Tables 2 and 3, although the standard errors are noticeably larger.

Table D.3: Table 2, raw citation counts

	1990–2015									2000–2015	
	All data			without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	11.422	13.031*	9.422	6.801	7.957	1.270	4.803	5.983	−0.666	11.535**	12.307**
	(7.457)	(7.446)	(7.752)	(7.463)	(7.426)	(7.852)	(7.435)	(7.412)	(7.775)	(4.859)	(5.244)
$\max t$		−2.502*	−2.627*		−4.710***	−4.979***		−4.560***	−4.835***		
		(1.314)	(1.357)		(1.722)	(1.752)		(1.685)	(1.715)		
$\max T$		3.243***	3.153***		4.883***	4.738***		4.794***	4.655***		
		(1.083)	(1.061)		(1.514)	(1.495)		(1.497)	(1.478)		
N			9.422			16.058***			16.009***		
			(6.306)			(5.334)			(5.307)		
$\sigma^2_M(\theta_M)$	70,164.653	70,083.640	70,075.796	33,420.300	33,281.697	33,218.399	33,109.244	32,983.929	32,921.026	7,095.112	7,673.547
$\sigma^2_F(\theta_F)$	5,254.691	5,252.894	5,246.777	5,378.518	5,373.619	5,373.440	4,920.204	4,919.264	4,918.826	4,277.620	5,359.791
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−13.762	−22.212	−28.677	−22.851	−28.573	−33.225	−19.432	−26.538	−32.426	−14.024	−0.631
	(27.923)	(29.137)	(30.499)	(27.762)	(29.046)	(30.493)	(28.137)	(28.542)	(30.436)	(25.170)	(24.392)
$\max t$		−7.149	−7.523*		−6.806	−7.040*		−9.340**	−9.648**		
		(4.430)	(4.349)		(4.279)	(4.245)		(4.001)	(3.969)		
$\max T$		7.997**	7.982**		7.229**	7.182**		9.716***	9.745***		
		(3.789)	(3.715)		(3.541)	(3.508)		(3.273)	(3.242)		
N			19.482*			13.726			14.098		
			(10.228)			(9.667)			(11.118)		
$\sigma^2_M(\theta_M)$	23,036	22,556	22,176	23,554	22,913	22,669	22,425	21,877	21,601	97,386	30,393
$\sigma^2_F(\theta_F)$	2,808	1,530	1,524	2,975	1,620	1,615	–	–	–	2,975	11,909
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (primary)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 2 except that the dependent variable is raw citation counts. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.4: Table 3, raw citation counts

	(1)	(2)	(3)
female co-author	10.813	71.921**	71.764***
	(29.147)	(28.447)	(26.669)
max t		−30.956***	−30.805***
		(4.854)	(4.660)
$\sigma^2_M(\theta_M)$	3,096	3,077	2,838
$\sigma^2_F(\theta_F)$	746	647	605
p-value (ratio)	0.000	0.000	0.000
No. obs.	314	314	314
Year×Journal	✓	✓	✓
Author	✓	✓	✓
JEL (primary)			✓

Note. Estimates are identical to those in Table 3 except that the dependent variable is raw citation counts. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.3 Controlling non-parametrically for the number of co-authors

Table D.5 replicates Table 2 but controls non-parametrically for the number of co-authors. (Given space constraints, we do not report the coefficients on each fixed effect for number of co-authors.) Results are very similar to those reported in Table 2.

Table D.5: Table 2, controlling non-parametrically for the number of co-authors

	1990–2015										
	All data			without *JEL* fixed effects			with *JEL* fixed effects			2000–2015	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	0.110***	0.125***	0.075**	0.117***	0.130***	0.066*	0.088**	0.101***	0.039	0.143***	0.150***
	(0.038)	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)	(0.038)	(0.038)	(0.039)	(0.040)	(0.041)
$\max t$		−0.010**	−0.011***		−0.013***	−0.015***		−0.014***	−0.016***		
		(0.004)	(0.004)		(0.004)	(0.004)		(0.004)	(0.004)		
$\max T$		0.020***	0.019***		0.023***	0.021***		0.023***	0.022***		
		(0.003)	(0.003)		(0.004)	(0.004)		(0.004)	(0.004)		
$\sigma_M^2(\theta_M)$	0.544	0.540	0.537	0.430	0.426	0.421	0.419	0.414	0.410	0.342	0.364
$\sigma_F^2(\theta_F)$	0.129	0.129	0.124	0.127	0.125	0.121	0.118	0.117	0.113	0.133	0.178
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−0.132	−0.235*	−0.316**	−0.191	−0.268**	−0.340**	−0.142	−0.225*	−0.321***	−0.237*	−0.158
	(0.133)	(0.134)	(0.136)	(0.128)	(0.131)	(0.133)	(0.121)	(0.121)	(0.123)	(0.136)	(0.137)
$\max t$		−0.012	−0.021		−0.010	−0.013		−0.020	−0.024		
		(0.022)	(0.021)		(0.021)	(0.020)		(0.021)	(0.020)		
$\max T$		0.036*	0.042**		0.029	0.028*		0.039**	0.039**		
		(0.020)	(0.019)		(0.018)	(0.017)		(0.018)	(0.016)		
$\sigma_M^2(\theta_M)$	0.858	0.811	0.793	0.871	0.820	0.806	0.826	0.782	0.767	1.528	1.077
$\sigma_F^2(\theta_F)$	0.153	0.117	0.113	0.162	0.124	0.119	–	–	–	0.162	0.277
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
N			✓			✓			✓		
JEL (primary)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 2 except that models in (3), (6) and (9) control non-parametrically for N. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.4 Controlling for institutional rank

In Tables D.6 and D.7, we control for institutional rank fixed effects. Results in both tables are very similar to the results reported in Tables 2 and 3, although the coefficient on female co-author in Table D.7 is more noisily estimated.

Table D.6: Table 2, controlling for institutional rank

	1990–2015									2000–2015	
	All data			without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	0.106***	0.122***	0.077**	0.115***	0.128***	0.068*	0.085**	0.098**	0.041	0.138***	0.144***
	(0.038)	(0.038)	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.041)	(0.041)
$\max t$		−0.010**	−0.011***		−0.012***	−0.015***		−0.013***	−0.015***		
		(0.004)	(0.004)		(0.004)	(0.004)		(0.004)	(0.004)		
$\max T$		0.020***	0.019***		0.022***	0.021***		0.023***	0.022***		
		(0.003)	(0.003)		(0.004)	(0.004)		(0.004)	(0.004)		
N			0.117***			0.142***			0.138***		
			(0.020)			(0.019)			(0.019)		
$\sigma^2_M(\theta_M)$	0.543	0.539	0.537	0.428	0.424	0.420	0.417	0.413	0.409	0.339	0.361
$\sigma^2_F(\theta_F)$	0.126	0.126	0.124	0.122	0.121	0.120	0.113	0.112	0.111	0.129	0.173
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−0.094	−0.203	−0.274**	−0.170	−0.252*	−0.314**	−0.138	−0.223*	−0.312**	−0.192	−0.121
	(0.138)	(0.137)	(0.134)	(0.133)	(0.135)	(0.132)	(0.126)	(0.125)	(0.124)	(0.141)	(0.136)
$\max t$		−0.003	−0.008		−0.006	−0.010		−0.019	−0.024		
		(0.022)	(0.021)		(0.020)	(0.020)		(0.020)	(0.020)		
$\max T$		0.030	0.030		0.027	0.027		0.037**	0.038**		
		(0.020)	(0.019)		(0.018)	(0.017)		(0.017)	(0.017)		
N			0.231***			0.201***			0.225***		
			(0.061)			(0.057)			(0.059)		
$\sigma^2_M(\theta_M)$	0.815	0.778	0.762	0.826	0.786	0.775	0.786	0.752	0.740	1.467	1.014
$\sigma^2_F(\theta_F)$	0.119	0.082	0.080	0.126	0.087	0.085	–	–	–	0.126	0.251
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Institution	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (primary)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 2 except that all models include fixed effects for the institutional rank of the author from the highest ranked institution. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.7: Table 3, controlling for institutional rank

	(1)	(2)	(3)
female co-author	−0.265	0.392	0.408
	(0.383)	(0.272)	(0.255)
$\max t$		−0.369***	−0.374***
		(0.072)	(0.069)
$\sigma^2_M(\theta_M)$	0.214	0.208	0.186
$\sigma^2_F(\theta_F)$	0.067	0.039	0.032
p-value (ratio)	0.000	0.000	0.000
No. obs.	314	314	314
Year×Journal	✓	✓	✓
Institution	✓	✓	✓
Author	✓	✓	✓
JEL (primary)			✓

Note. Estimates are identical to those in Table 3 except that all models include fixed effects for the institutional rank of the author from the highest ranked institution. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.5 Controlling for secondary and tertiary *JEL* codes

In Table D.8, we replicate columns (7)–(9) of Table 2 controlling for secondary (columns (1)–(3)) and tertiary *JEL* categories (columns (4)–(6)), although the latter only for male authors.[3] Results are roughly similar to those reported in Table 2.

Table D.9 similarly replicates column (3) of Table 3. The coefficient on female co-author is about 20–30 percent higher after conditioning on secondary or tertiary *JEL* codes.

Table D.8: Table 2, controlling for secondary and tertiary *JEL* codes

	Secondary *JEL* fixed effects			Tertiary *JEL* fixed effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Returns to men from co-authoring with women						
female co-author(s)	0.072*	0.085**	0.021	0.088**	0.102**	0.049
	(0.040)	(0.040)	(0.040)	(0.042)	(0.042)	(0.043)
$\max t$		−0.010**	−0.013***		−0.011**	−0.014***
		(0.004)	(0.004)		(0.004)	(0.004)
$\max T$		0.021***	0.019***		0.022***	0.021***
		(0.004)	(0.004)		(0.004)	(0.004)
N			0.151***			0.141***
			(0.019)			(0.022)
$\sigma^2_M(\theta_M)$	0.395	0.392	0.388	0.309	0.306	0.304
$\sigma^2_F(\theta_F)$	0.070	0.069	0.066	–	–	–
p-value (ratio)	0.000	0.000	0.000	–	–	–
No. obs.	9,465	9,465	9,465	9,465	9,465	9,465
Returns to women from co-authoring with men						
male co-author(s)	−0.112	−0.216	−0.266*			
	(0.146)	(0.142)	(0.144)			
$\max t$		−0.012	−0.012			
		(0.024)	(0.024)			
$\max T$		0.041**	0.038*			
		(0.019)	(0.019)			
N			0.123*			
			(0.067)			
$\sigma^2_M(\theta_M)$	0.702	0.660	0.652			
$\sigma^2_F(\theta_F)$	–	–	–			
p-value (ratio)	–	–	–			
No. obs.	1,099	1,099	1,099			
Year×Journal	✓	✓	✓	✓	✓	✓
Author	✓	✓	✓	✓	✓	✓
JEL (secondary)	✓	✓	✓			
JEL (tertiary)				✓	✓	✓

Note. Estimates are identical to those in columns (7)–(9) of Table 2 except that the first three columns replace fixed effects for primary *JEL* categories with fixed effects for secondary *JEL* categories and the last three columns replace them with fixed effects for tertiary *JEL* categories. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

[3] The number of female authors with two or more co-authored papers is too small to reliably estimate Equation (2) in columns (4)–(6). For similar reasons, we also do not estimate $\sigma_i^2(\theta_{iF})$ in the sample of male authors when conditioning on tertiary *JEL* codes.

Table D.9: Table 3, controlling for secondary and tertiary *JEL* codes

	(1)	(2)
female co-author	0.687***	0.751***
	(0.174)	(0.164)
max t	−0.187***	−0.157***
	(0.025)	(0.029)
$\sigma^2_M(\theta_M)$	0.190	0.093
$\sigma^2_F(\theta_F)$	0.050	0.010
p-value (ratio)	0.000	0.000
No. obs.	393	447
Year×Journal	✓	✓
Author	✓	✓
JEL (secondary)	✓	
JEL (tertiary)		✓

Note. Estimates are identical to those column (3) of Table 3 except that column (1) replaces primary *JEL* fixed effects with secondary *JEL* fixed effects and in column (2) they are replaced with tertiary *JEL* fixed effects. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.6 Controlling for the length of the bibliography

In Tables D.10 and D.11 we replicate Tables 2 and 3, respectively, controlling for the number of research outputs each paper cites in its bibliography. (See Appendix C.8 for a discussion and justification of this control variable.)

The first panel of Table D.10 is almost identical to the first panel of Table 2 and Table D.11 is almost identical to Table 3. Among female authors, the direction of the coefficient on male co-authors is consistently negative in Table D.10, although the magnitude declines somewhat relative to the corresponding results in Table 2.

In both the male and female samples, the length of the bibliography is positively associated with the number of citations an article receives, similar to what we found in Appendix C.8. Interestingly, however, the coefficient on bibliography length is a tightly estimated zero in Table D.11. This suggests that there may not be a relationship between bibliography length and number of citations received after controlling very carefully for author-specific qualities.

Table D.10: Table 2, controlling for the length of the bibliography

	1990–2015									2000–2015	
	All data			without *JEL* fixed effects			with *JEL* fixed effects				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Returns to men from co-authoring with women											
female co-author(s)	0.110***	0.125***	0.080**	0.121***	0.133***	0.073*	0.096**	0.108***	0.051	0.143***	0.146***
	(0.037)	(0.037)	(0.037)	(0.038)	(0.038)	(0.039)	(0.038)	(0.038)	(0.038)	(0.040)	(0.040)
max t		−0.009**	−0.010**		−0.011***	−0.013***		−0.012***	−0.014***		
		(0.004)	(0.004)		(0.004)	(0.004)		(0.004)	(0.004)		
max T		0.019***	0.018***		0.021***	0.019***		0.022***	0.020***		
		(0.003)	(0.003)		(0.003)	(0.003)		(0.004)	(0.003)		
N			0.117***			0.144***			0.139***		
			(0.019)			(0.019)			(0.018)		
bibl. length	0.019***	0.019***	0.019***	0.017***	0.016***	0.017***	0.017***	0.017***	0.017***	0.015***	0.016***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\sigma^2_M(\theta_M)$	0.508	0.505	0.503	0.398	0.395	0.392	0.387	0.384	0.380	0.312	0.333
$\sigma^2_F(\theta_F)$	0.122	0.121	0.120	0.121	0.119	0.118	0.114	0.112	0.111	0.127	0.171
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
No. obs.	13,060	13,060	13,060	9,465	9,465	9,465	9,465	9,465	9,465	7,009	7,009
Returns to women from co-authoring with men											
male co-author(s)	−0.038	−0.146	−0.226*	−0.116	−0.196	−0.266**	−0.073	−0.162	−0.262**	−0.170	−0.104
	(0.136)	(0.139)	(0.137)	(0.128)	(0.133)	(0.133)	(0.119)	(0.121)	(0.123)	(0.138)	(0.143)
max t		−0.008	−0.012		−0.006	−0.010		−0.016	−0.021		
		(0.021)	(0.020)		(0.021)	(0.020)		(0.021)	(0.020)		
max T		0.034*	0.034*		0.027	0.027		0.038**	0.038**		
		(0.020)	(0.018)		(0.019)	(0.018)		(0.018)	(0.017)		
N			0.251***			0.226***			0.255***		
			(0.057)			(0.056)			(0.058)		
bibl. length	0.015***	0.016***	0.016***	0.011***	0.011***	0.012***	0.011***	0.011***	0.012***	0.011***	0.009**
	(0.004)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
$\sigma^2_M(\theta_M)$	0.826	0.778	0.762	0.840	0.789	0.776	0.796	0.752	0.740	1.432	1.036
$\sigma^2_F(\theta_F)$	0.148	0.106	0.103	0.157	0.113	0.109	–	–	–	0.157	0.271
p-value (ratio)	0.000	0.000	0.000	0.000	0.000	0.000	–	–	–	0.000	0.000
No. obs.	1,230	1,230	1,230	1,099	1,099	1,099	1,099	1,099	1,099	926	926
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Author	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
JEL (primary)							✓	✓	✓		
Year											✓

Note. Estimates are identical to those in Table 2, except that all models control for the length of a paper's bibliography. (See Appendix B for information on how this indicator was constructed.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

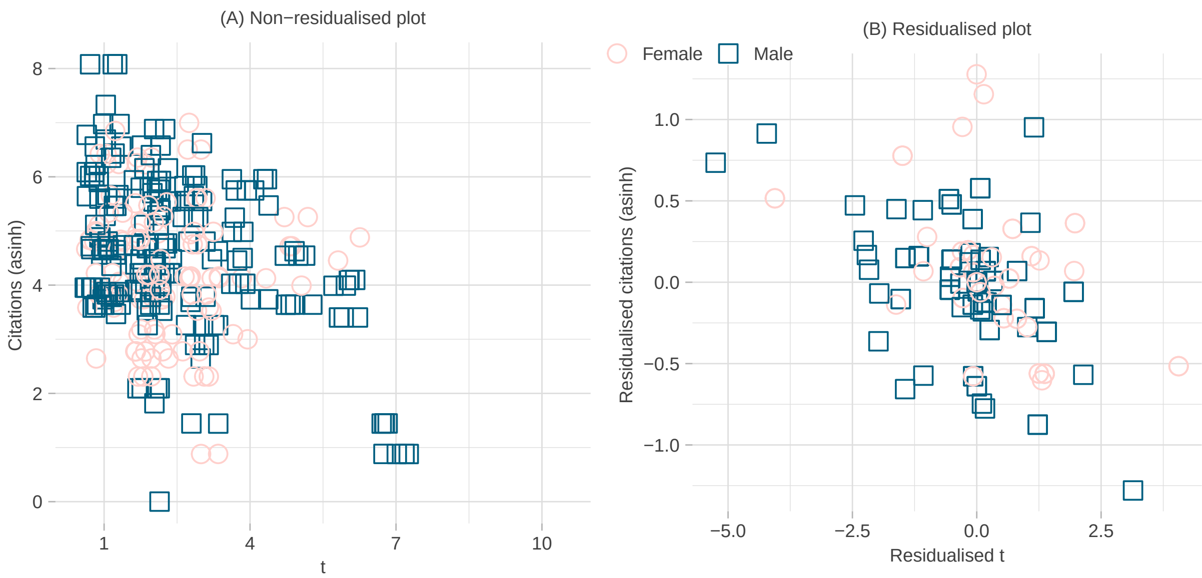
Table D.11: Table 3, controlling for the length of the bibliography

	(1)	(2)	(3)
female co-author	0.235	0.614***	0.610***
	(0.228)	(0.226)	(0.211)
max t		−0.175***	−0.176***
		(0.031)	(0.029)
bibl. length	0.004	0.000	−0.001
	(0.005)	(0.005)	(0.005)
$\sigma_M^2(\theta_M)$	0.290	0.288	0.257
$\sigma_F^2(\theta_F)$	0.150	0.114	0.103
p-value (ratio)	0.000	0.000	0.000
No. obs.	314	314	314
Year×Journal	✓	✓	✓
Author	✓	✓	✓
JEL (primary)			✓

Note.
Estimates are identical to those in Table 3, except that all models control for the length of a paper's bibliography. (See Appendix B for information on how this indicator was constructed.) ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D.7 Table 3, covariate balance

By design, the sample of senior authors used to estimate Table 3 fixes N and $\max T$, conditional on author. For each author, however, t varies over time and appears somewhat imbalanced between treatment and control groups, particularly after accounting for author-specific fixed effects and journal-year interaction dummies (Figure D.1)—*i.e.*, conditional on author, year and journal, the senior men in our sample were slightly more experienced when they co-authored with junior women than they were when they co-authored with junior men. For that reason, we additionally control for $\max t$ in columns (2)–(4) of Table 3.



Note. Graph (A) plots $\max t$ (x-axis) against asinh-transformed citations (y-axis) by co-author sex for the sample of senior male authors satisfying the conditions outlined in Section 5.2. Graph (B) plots the residuals of both variables after accounting for author-specific fixed effects and journal-year interaction dummies.

Figure D.1: $\max t$ balance among senior men

D.8 Table 3, list of senior men

Table D.12: Table 3, list of senior men

Daron Acemoglu	Mark Gertler	Hervé Moulin
Alberto Alesina	Robert E. Hall	Ulrich K. Muller
James Andreoni	James D. Hamilton	Thomas R. Palfrey
Donald W. K. Andrews	Yongmiao Hong	Martin Pesendorfer
Robert J. Barro	Hugo A. Hopenhayn	Peter C. B. Phillips
Robert B. Barsky	Joel L. Horowitz	Charles R. Plott
B. Douglas Bernheim	Hanan G. Jacoby	Debraj Ray
Michele Boldrin	Boyan Jovanovic	Diego Restuccia
George J. Borjas	Edi Karni	Jean-Marc Robin
Stephen G. Bronars	Brian Knight	Andrés Rodríguez-Clare
Martin J. Browning	Michael Kremer	Alvin E. Roth
Pierre-André Chiappori	Pravin Krishna	Ariel Rubinstein
John H. Cochrane	Alan B. Krueger	Lones Smith
Timothy Cogley	Peter Kuhn	Joel Waldfogel
Vincent P. Crawford	Gary D. Libecap	Jörgen W. Weibull
Raymond J. Deneckere	Steven A. Matthews	David E. Weinstein
Gregory K. Dow	Paul R. Milgrom	Halbert White
John Duffy	Espen R. Moen	Randall Wright
Christopher J. Flinn	John Morgan	

D.9 Reconciling Table 3 with Card *et al.* (2020)

Card *et al.* (2020) do not find a difference in citations between mixed-gendered papers with a senior male co-author relative to papers co-authored by all-male teams. In contrast, the evidence presented in Table 3 suggests that papers by senior male authors are cited more when they are co-authored with junior women compared to junior men. We believe these differing results are due to co-author composition effects that our within-author analysis is better able to account for.

To illustrate what we mean, Table D.13 displays results from a regression of $\max T$ on female, N and $\max t$ in the sample of co-authored articles where the senior author was male.[4] These results suggest that when female authors co-author top-five papers with senior men, the reputation of those senior men (as captured by $\max T$) is *lower* than the reputation of the senior men who co-author entirely with other men. Thus, the Matthew effect likely skews citations to papers by all-male teams more than it skews citations to mixed-gendered papers with a senior male co-author, conditional on quality. As a result, a between-paper analysis—as conducted by Card *et al.* (2020)—could conclude that mixed-gendered papers with a senior male co-author are not cited more than papers co-authored by all-male teams, *even though* quality is, on average, higher in the former than it is in the latter.

In Table 3, we fix the seniority of the senior male co-author. As a result, our analysis is better able to hold the Matthew effect constant between "treated" (*i.e.*, senior male authors co-authoring with junior women) and "control" groups (*i.e.*, those same senior men co-authoring with junior men).

Table D.13: Relationship between $\max T$ and the gender of junior co-authors

	(1)	(2)	(3)	(4)	(5)	(6)
female	−0.580	−1.134***	−1.033***			
	(0.481)	(0.238)	(0.227)			
1+ female				0.486	−0.732***	−0.542***
				(0.392)	(0.174)	(0.166)
N		−0.191	−0.171		−0.046	0.025
		(0.124)	(0.114)		(0.112)	(0.106)
$\max t$		1.272***	1.217***		1.268***	1.217***
		(0.015)	(0.015)		(0.015)	(0.014)
No. obs.	5,349	5,349	3,705	5,645	5,645	3,984
R^2	0.097	0.718	0.814	0.091	0.725	0.818
Year×Journal	✓	✓	✓	✓	✓	✓
JEL (primary)			✓			✓

Note. OLS regression of $\max T$ on female (defined as 50% or more female co-authors in columns (1)–(3) (mixed-gendered papers with fewer than 50% female authors are dropped) and at least one female co-author in columns (4)–(6)). Sample restricted to papers co-authored by two or more authors, where the senior author—defined as having the most top-five publications at the time the paper was published—was male. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

[4] Papers are assumed to be co-authored by a senior man if the co-author with the most top-five publications at the time of publication was a man. (Co-authored papers with a senior female co-author are dropped.)

E Right-tail confounders

In order to illustrate how gender differences in raw citation counts at the mean may be distorted by a small number of extremely famous—and disproportionately male—economists, we control for "superstar" (Appendix E.1) and Nobel Prize winning authors (Appendix E.2).

E.1 Superstar authors

We define "superstars" as authors who satisfy one or more of the following criteria:

1. 17 or more top-five publications (one percent of all authors);
2. 10 or more top-five publications, one of which is cited at least 2,500 times (0.2 percent of all authors);
3. 5 or more top-five publications, one of which is cited at least 5,000 times (0.1 percent of all authors).

The first criteria defines superstar according to quantity, alone. It is set as one plus the lifetime number of publications of the most prolific female economist as of December 2015 (Esther Duflo). Criteria two and three account for famous economists who are less prolific—*e.g.*, Paul Krugman—operate in fields with slower production functions—*e.g.*, industrial organisation—or publish extensively in other disciplines—*e.g.*, Daniel Kahneman. General results and conclusions do not change by making marginal adjustments to any criteria—including redefining condition (1) to include every male and female author with at least 10–15 publications.

1.2 percent of authors satisfy at least one condition. On average, each has published 21 times in a top-five journal; his highest cited paper is cited 1,844 times. Almost a third either won the Nobel Prize, the John Bates Clark medal or both. All are male. See Table E.1 for a list of their names.

E.1.1 Results

Tables E.2–E.4 illustrate the effect of super-stardom on gender differences in raw citation counts using articles as the unit of analysis. Table E.2 is estimated using all observations. Column (1) controls only for journal-year fixed effects and the female composition of a paper. It suggests that male-authored papers receive, on average, about 10 more citations than female-authored papers. The sign on the coefficient substantially declines, however, after including the superstar dummy (column (2)) and then flips (but is insignificant) after adding fixed effects for each superstar author (column (3)). Columns (4)–(9) control for N, $\max t$ and $\max T$. The coefficient on female generally hovers around zero, but jumps to 7 citations in the final column.

Older male-authored papers likely drive the bulk of superstar bias. Their impact, however, should attenuate the closer an article is to its date of publication. Tables E.3 and E.4 support this hypothesis. They reproduce results from Table E.2, but restrict the sample to papers published after 1990 and 2000, respectively. The coefficients on female in Table E.3 are universally larger than corresponding figures from Table E.2; the estimate in the final column suggests female-authored papers receive, on average, 9 more citations than male-authored papers after controlling for journal-year fixed effects, N, $\max t$, $\max T$ and superstar author fixed effects. When data are restricted to articles published after 2000, female-authored papers are consistently cited more frequently than male-authored papers. Moreover, controlling for super-stardom has much less of an impact on the observed relationship between the female composition of a paper and its citations (Table E.4).

Table E.1: List of superstar men

Abel, Andrew B.	Fisher, Franklin M.	Pakes, Ariel
Acemoglu, Daron	Fudenberg, Drew	Palfrey, Thomas R.
Aghion, Philippe	Gale, Douglas	Persson, Torsten
Alesina, Alberto	Granger, Clive W. J.	Phillips, Peter C. B.
Andrews, Donald W. K.	Green, Jerry R.	Plott, Charles R.
Arellano, Manuel	Grossman, Gene M.	Postlewaite, Andrew
Banerjee, Abhijit V.	Grossman, Sanford J.	Ray, Debraj
Barro, Robert J.	Gruber, Jonathan	Robinson, Peter M.
Becker, Gary S.	Gul, Faruk	Romer, David H.
Bénabou, Roland	Hamilton, James D.	Rosen, Sherwin
Bernheim, B. Douglas	Hansen, Lars Peter	Rosenzweig, Mark R.
Besley, Timothy J.	Hart, Oliver D.	Roth, Alvin E.
Blackorby, Charles	Hausman, Jerry A.	Rubinstein, Ariel
Blanchard, Olivier J.	Heckman, James J.	Saez, Emmanuel
Blundell, Richard W.	Helpman, Elhanan	Samuelson, Larry
Bolton, Patrick	Jackson, Matthew O.	Sargent, Thomas J.
Browning, Martin J.	Jovanovic, Boyan	Scheinkman, José A.
Caballero, Ricardo J.	Kahneman, Daniel	Shleifer, Andrei
Campbell, John Y.	Kehoe, Patrick J.	Stein, Jeremy C.
Caplin, Andrew S.	Kremer, Michael	Stiglitz, Joseph E.
Card, David E.	Krugman, Paul R.	Tirole, Jean
Chiappori, Pierre-André	Laffont, Jean-Jacques	Tversky, Amos
Cooper, Russell	Laroque, Guy	Vishny, Robert W.
Crawford, Vincent P.	Levine, David K.	Weil, David N.
Deaton, Angus S.	Levitt, Steven D.	Weitzman, Martin L.
Diamond, Peter A.	List, John A.	White, Halbert
Dixit, Avinash K.	Mankiw, N. Gregory	Wolpin, Kenneth I.
Engle, Robert F.	Maskin, Eric S.	Wright, Randall
Epstein, Larry G.	Milgrom, Paul R.	Zame, William R.
Fehr, Ernst	Murphy, Kevin M.	
Feldstein, Martin S.	Newey, Whitney K.	

Table E.2: The impact of super-stardom (1950–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	−9.943*	0.233	3.217	−8.536*	0.755	3.495	−0.124	−0.531	7.133
	(5.124)	(5.098)	(5.07)	(5.087)	(5.083)	(5.047)	(5.046)	(5.134)	(4.976)
superstar		107.874***			103.993***			94.906***	
		(17.017)			(17.132)			(34.379)	
N				21.019***	13.188***	9.331**	16.268***	16.287***	7.248*
				(4.384)	(4.338)	(4.224)	(4.5)	(4.504)	(4.171)
$\max t$							−6.817***	−6.933***	−6.748***
							(1.118)	(1.124)	(1.065)
$\max T$							6.796***	3.878***	8.641***
							(0.913)	(1.275)	(0.826)
No. obs.	10,566	10,553	10,553	10,566	10,553	10,553	10,566	10,553	10,553
R^2	0.056	0.069	0.164	0.058	0.070	0.164	0.070	0.074	0.173
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Superstar authors			✓			✓			✓

Note. Figures correspond to coefficients from an OLS regression of raw citation counts on a dummy variable equal to 1 if the paper was authored by at least 50 percent women and 0 otherwise. (Mixed gendered papers with fewer than 50 percent female authors are dropped.) Superstar is a binary variable equal to 1 if at least one author on a paper satisfies the criteria defined in Appendix E.1. Superstar fixed effects account for each superstar author. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.3: The impact of super-stardom (1990–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	−3.858	2.103	4.638	−1.458	3.441	5.689	6.059	5.926	9.414*
	(5.708)	(5.464)	(5.515)	(5.65)	(5.456)	(5.513)	(5.583)	(5.64)	(5.503)
superstar		63.788***			57.675***			26.239	
		(11.584)			(11.309)			(22.708)	
N				22.555***	17.812***	16.613***	18.689***	18.624***	15.965***
				(4.28)	(4.045)	(3.369)	(4.163)	(4.158)	(3.497)
$\max t$							−7.026***	−7.173***	−8.394***
							(1.166)	(1.179)	(1.114)
$\max T$							7.039***	6.322***	9.209***
							(1.005)	(1.165)	(1.012)
No. obs.	6,129	6,124	6,124	6,129	6,124	6,124	6,129	6,124	6,124
R^2	0.105	0.116	0.191	0.111	0.119	0.194	0.131	0.132	0.209
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Superstar authors			✓			✓			✓

Note. Columns display estimates identical to those in Table E.2 except that only articles published after 1990 are included. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.4: The impact of super-stardom (2000–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	8.942	9.980*	11.427**	10.636*	11.221**	12.583**	13.671**	13.894**	14.662***
	(5.483)	(5.423)	(5.522)	(5.501)	(5.445)	(5.546)	(5.498)	(5.506)	(5.676)
superstar		14.358**			9.547			−25.385***	
		(6.286)			(6.21)			(8.486)	
N				13.093***	12.366***	11.981***	11.208***	11.127***	10.435***
				(2.469)	(2.435)	(2.447)	(2.368)	(2.366)	(2.551)
$\max t$							−4.824**	−4.533**	−4.729***
							(1.98)	(1.955)	(1.821)
$\max T$							4.780***	5.326***	5.930***
							(1.771)	(1.828)	(1.49)
No. obs.	3,982	3,979	3,979	3,982	3,979	3,979	3,982	3,979	3,979
R^2	0.182	0.184	0.211	0.189	0.190	0.216	0.206	0.208	0.227
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Superstar authors			✓			✓			✓

Note. Columns display estimates identical to those in Table E.2 except that only articles published after 2000 are included. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

E.2 Nobel Prize-winning authors

In this Appendix, we swap our *ad hoc* definition of "superstar" (Appendix E.1) with fixed effects (and a binary variable) for authors who had won the Nobel Prize before the citation data were last updated.

About 0.9 percent of authors in our data are Nobel Prize winners. (See Table E.5 for a list of their names.) On average, each has published 10 papers in a top-five journal. Their highest cited paper was cited, on average, 1,515 times.

E.2.1 Results

Results in Tables E.6, E.7 and E.8 closely mirror corresponding results from Appendix E.1. Controlling for Nobel Prize winners reduces the magnitude of the coefficient on female authorship (Table E.6) but the change is less pronounced when the sample is restricted to later years (Tables E.7 and E.8). Among articles published after 2000 (Table E.8), female-authored papers receive, on average, 9–14 more citations compared to male-authored papers and accounting for Nobel Prize winners does not observably impact this gap.

Table E.5: List of Nobel Prize winners

Akerlof, George A.	Koopmans, Tjalling C.	Samuelson, Paul A.
Allais, Maurice	Krugman, Paul R.	Sargent, Thomas J.
Arrow, Kenneth J.	Kydland, Finn E.	Scholes, Myron S.
Aumann, Robert J.	Lucas, Robert E. (Jr.)	Schultz, Theodore W.
Becker, Gary S.	Markowitz, Harry M.	Selten, Reinhard
Buchanan, James M.	Maskin, Eric S.	Sen, Amartya K.
Deaton, Angus S.	McFadden, Daniel L.	Shapley, Lloyd S.
Debreu, Gerard	Merton, Robert C.	Shiller, Robert J.
Diamond, Peter A.	Miller, Merton H.	Simon, Herbert A.
Engle, Robert F.	Mirrlees, James A.	Sims, Christopher A.
Fama, Eugene F.	Modigliani, Franco	Smith, Vernon L.
Friedman, Milton	Mortensen, Dale T.	Solow, Robert M.
Frisch, Ragnar	Mundell, Robert A.	Spence, A. Michael
Granger, Clive W. J.	Myerson, Roger B.	Stigler, George J.
Hansen, Lars Peter	Nordhaus, William D.	Stiglitz, Joseph E.
Harsanyi, John C.	North, Douglass C.	Stone, Richard
Hart, Oliver D.	Ostrom, Elinor	Thaler, Richard H.
Heckman, James J.	Phelps, Edmund S.	Tinbergen, Jan
Holmström, Bengt	Pissarides, Christopher A.	Tirole, Jean
Hurwicz, Leonid	Prescott, Edward C.	Tobin, James
Kahneman, Daniel	Romer, Paul M.	Williamson, Oliver E.
Klein, Lawrence R.	Roth, Alvin E.	

Table E.6: The impact of the Nobel Prize (1950–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	−9.943*	−4.343	−1.722	−8.536*	−3.359	−0.963	−0.124	−0.615	5.773
	(5.124)	(5.19)	(5.133)	(5.087)	(5.165)	(5.088)	(5.046)	(5.098)	(4.99)
Nobel		218.171***			214.573***			197.538***	
		(39.49)			(39.428)			(43.683)	
N				21.019***	16.088***	13.363***	16.268***	16.143***	9.547**
				(4.384)	(4.229)	(4.069)	(4.5)	(4.486)	(3.932)
$\max t$							−6.817***	−7.489***	−7.046***
							(1.118)	(1.164)	(1.038)
$\max T$							6.796***	5.415***	6.941***
							(0.913)	(0.92)	(0.784)
No. obs.	10,566	10,566	10,566	10,566	10,566	10,566	10,566	10,566	10,566
R^2	0.056	0.077	0.193	0.058	0.079	0.194	0.070	0.086	0.205
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nobel authors			✓			✓			✓

Note. Figures correspond to coefficients from an OLS regression of raw citation counts on a dummy variable equal to 1 if the paper was authored by at least 50 percent women and 0 otherwise. (Mixed gendered papers with fewer than 50 percent female authors are dropped.) Nobel is a dummy variable equal to 1 if at least one author on a paper is a Nobel Prize winner; Nobel fixed effects account for each Nobel Prize winning author. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.7: The impact of the Nobel Prize (1990–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	−3.858	−2.240	−2.013	−1.458	−0.033	0.133	6.059	5.932	7.272
	(5.708)	(5.71)	(5.62)	(5.65)	(5.646)	(5.554)	(5.583)	(5.576)	(5.461)
Nobel		81.815***			77.394***			64.648***	
		(19.865)			(20.094)			(23.118)	
N				22.555***	21.564***	22.076***	18.689***	19.257***	17.999***
				(4.28)	(4.345)	(4.353)	(4.163)	(4.13)	(4.114)
$\max t$							−7.026***	−7.698***	−7.586***
							(1.166)	(1.147)	(1.149)
$\max T$							7.039***	7.011***	7.625***
							(1.005)	(1.011)	(1.06)
No. obs.	6,129	6,129	6,129	6,129	6,129	6,129	6,129	6,129	6,129
R^2	0.105	0.110	0.135	0.111	0.115	0.140	0.131	0.134	0.162
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nobel authors			✓			✓			✓

Note. Columns display estimates identical to those in Table E.6 except that only articles published after 1990 are included. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.8: The impact of the Nobel Prize (2000–2015)

	Model 1			Model 2			Model 3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	8.942	9.408*	8.164	10.636*	11.042**	9.834*	13.671**	13.754**	12.492**
	(5.483)	(5.513)	(5.382)	(5.501)	(5.529)	(5.396)	(5.498)	(5.504)	(5.381)
Nobel		22.644*			20.554*			18.857	
		(12.373)			(12.472)			(12.854)	
N				13.093***	12.957***	13.404***	11.208***	11.419***	11.684***
				(2.469)	(2.48)	(2.532)	(2.368)	(2.364)	(2.389)
$\max t$							−4.824**	−5.062**	−5.165***
							(1.98)	(1.983)	(1.991)
$\max T$							4.780***	4.845***	4.957***
							(1.771)	(1.772)	(1.791)
No. obs.	3,982	3,982	3,982	3,982	3,982	3,982	3,982	3,982	3,982
R^2	0.182	0.183	0.196	0.189	0.189	0.203	0.206	0.206	0.220
Year×Journal	✓	✓	✓	✓	✓	✓	✓	✓	✓
Nobel authors			✓			✓			✓

Note. Columns display estimates identical to those in Table E.6 except that only articles published after 2000 are included. ***, ** and * statistically significant at 1%, 5% and 10%, respectively. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

References

Card, D. *et al.* (2020). "Are referees and editors in economics gender neutral?" *Quarterly Journal of Economics* 135 (1), pp. 269–327.

Cherrier, B. (2017). "Classifying economics: a history of *JEL* codes". *Journal of Economic Literature* 55 (2), pp. 545–579.

Hajeck, B. (2015). *Random Processes for Engineers*. Cambridge: Cambridge University Press.

Hengel, E. (2017). "Publishing while female. Are women held to higher standards? Evidence from peer review." Cambridge Working Paper Economics: 1753.

– (2022). "Publishing while female. Are women held to higher standards? Evidence from peer review." *The Economic Journal* 132 (648), pp. 2951–2991.

Hengel, E. and E. Moon (2020). "Gender and quality at top economics journals". Mimeo.

Lewis, T. and J. W. Thompson (1981). "Dispersive distributions, and the connection between dispersivity and strong unimodality". *Journal of Applied Probability* 18 (1), pp. 76–90.

Shaked, M. (1982). "Dispersive ordering of distributions". *Journal of Applied Probability* 19 (2), pp. 310–320.

Waltman, L. (2016). "A review of the literature on citation impact indicators". *Journalof Informetrics* 10, pp. 365–391.

Web of Science (2018). *Web of Science Core Collection*.

Zitt, M. and H. Small (2008). "Modifying the journal impact factor by fractional citation weighting: the audience factor". *Journal of the American Society for Information Science and Technology* 59 (11), pp. 1856–1860.