Gender and the time cost of peer review*

Diane Alexander† Olga Gorelkina‡ Erin Hengel§

December 2021

Abstract

In this paper, we investigate one factor that can directly contribute to—as well as indirectly shed light on the other causes of—the gender gap in academic publications: time spent in peer review. To study our problem, we link administrative data from an economics field journal with bibliographic and demographic information on the articles and authors it publishes. Our results suggest that in each round of review, referees spend 4.4 more days reviewing female-authored papers and female authors spend 12.3 more days revising their manuscripts. However, both gender gaps decline—and eventually disappear—as the same referee reviews more papers. This pattern suggests novice referees initially statistically discriminate against female authors; as their information about and confidence in the refereeing process improves, however, the gender gaps fall.

*We are especially grateful to Richard Tol for substantial support and assistance and to Amanda McFarland for excellent research assistance. We also thank Sarah Smith and seminar participants at the University of Liverpool and American University for useful comments and suggestions.

†The Wharton School, dianeealexand@gmail.com.

‡University of Liverpool, ogorelkina@gmail.com.

§University College London, erin.hengel@gmail.com.

1 Introduction

On average, men publish more papers than women (West *et al.* 2012). The gap remains large after controlling for observable factors (see, *e.g.*, Ductor *et al.* 2018). It also emerges surprisingly early—women take five years to reach the same annual rate of output that men achieve in two years (Symonds *et al.* 2006).

Yet this gender "productivity gap" is not identical across field nor is it universally present in all output dimensions relevant to academia. For example, the gap is greatest for publications in top journals but declines where research is less expensive and disappears for papers published in non-top journals and books (Duch *et al.* 2012; Mayer and Rathmann 2018). Survey evidence also suggests that it reverses direction when productivity is measured in terms of teaching and service to the profession or department (Aldercotte *et al.* 2017; Guarino and Borden 2017).

In this paper, we investigate one specific factor that can directly contribute to—as well as indirectly shed light on the other causes of—the gender publication gap: time spent in peer review. Our analysis focuses on the following three questions: (i) do referees review papers by female authors as quickly as they review papers by male authors? (ii) do women spend more time responding to referees? and (iii) if the answers to (i) and/or (ii) are "yes", to what extent do these gaps depend on how informed referees are about a journal's standards of acceptance and their skill and accuracy at assessing manuscript quality?

To study our problem, we link administrative data from the journal *Energy Economics* with bibliographic and demographic information on the articles and authors it publishes. Our administrative dataset contains comprehensive information on every round of the review process for accepted articles—including dates authors submitted (or re-submitted) their papers, dates referees agreed to review and eventually returned their reports, and unique identifiers that track referees and authors as they review or submit multiple manuscripts. We augment these data with citations, secondary *JEL* codes and the genders and institutions of corresponding authors.

Our results suggest that in each round of review, referees spend 4.4 more days reviewing female-authored papers and female authors spend 12.3 more days revising their manuscripts, conditional on proxies for quality (citations, referees' recommendations and editors' decisions) and fixed effects for year of submission. Both gaps remain constant or increase after including fixed effects for editors, referees, secondary *JEL* codes and authors' institutions. The are also robust to controlling for author prominence, manuscript length and number of co-authors.

Assuming the allocation of referees across papers is independent of author gender conditional on controls and at least one of our proxies for manuscript quality is not biased in favour of women, we argue that referees spend more time evaluating female-authored papers due to some combination of taste-based, statistical and/or institutional discrimination. For example, referees could delay reviewing women's papers because they are more reluctant to advance their careers (taste-based discrimination). They may also be more careful because they believe—correctly or incorrectly—that women submit lower quality work (statistical discrimination). Alternatively, current refereeing norms and practices may have evolved in a way that disadvantages women (institutional discrimination)—for example, referees may attend seminars and conferences to learn about the papers they review, but childcare responsibilities could make women less able to present at these events.

In contrast, the gender gap in revision times could result from discriminatory *or* non-discriminatory factors (or both). For example, prejudicial preferences combined with institutional discrimination can mean referees respect women's time less and/or aren't as concerned about the negative externalities low quality reports impose on female authors. Referees may also ask for more changes to female-authored papers because they believe—again, correctly or incorrectly—that their quality is lower (statistical discrimination). Among non-discriminatory factors, greater childcare obligations, teaching loads or other service responsibilities may force women to push their eventual resubmission dates further into the future. Women may also spend more time addressing referee concerns because they are less informed about the appropriate way to interpret their reports.

A common factor possibly driving both gender gaps is incomplete information—especially on the part of referees. Better informed referees will be more knowledgeable about whether a paper meets the standards of acceptance at *Energy Economics*; the signal they receive about its quality will also be more accurate—and contain fewer gender-specific distortions—than the signal received by less informed referees. Clearer signals should also boost one's ability to write less ambiguous, easier-to-implement reports, all else equal.

To proxy for how informed a referee is, we use his previous experience reviewing for *Energy Economics*. First, experienced referees are undoubtedly more knowledgeable about *Energy Economics*'s standards of acceptance—indeed, by frequently refereeing for it, they regularly set those standards. Second, thanks to a combination of learning-by-doing and actively defining acceptance standards, experienced referees are probably better at distinguishing low from high quality papers compared to their less experienced selves. Both factors likely also improve their ability to specify the changes that would need to be made before a paper should be accepted (again, all else equal).[1]

We use this insight—as well as the assumption that referee assignment across author gender is orthogonal to the former's experience—to identify the impact referees' incomplete information has on gender gaps in reviewing and revision times. According to our evidence, it plays a pivotal role. The gender gap in time spent revising is greater in papers that are reviewed by novice referees than in papers reviewed by expert referees, both before and after accounting for referee fixed effects. This pattern suggests that novice referees' incomplete information causes them to write reports that female authors find especially difficult to respond to—either because women are more careful about replying to uncertain, vague or unclear requests (*e.g.*, because they are more risk averse) or because novice referees write them tougher reports (statistical discrimination).

A similar analysis of time spend with referees supports the latter hypothesis. Novice and moderately experienced referees spend more time evaluating female-authored papers than they do evaluating male-authored papers; among highly experienced referees, however, the gap disappears or reverses. (Both conclusions hold with and without referee fixed effects.) This pattern suggests novice referees initially statistically discriminate against female authors. As they gain experience, however, their ability to identify quality in men's and women's papers converges, so the gap in reviewing time declines.

This paper makes several contributions. First, we join a large literature empirically linking imperfect information to biased outcomes (see *e.g.*, Benson *et al.* 2021; Bohren *et al.* 2019; Sarsons 2019). We add to this work by showing that incomplete information can have discriminatory consequences even when the decisions themselves appear bias-free. As a result, successful attempts by decision-makers to overcome bias probably still impose costs on the individuals they are biased against—for example, in our case women must wait longer for reports and spend more time revising.

Our second contribution is to propose a solution. As referees gain knowledge about the process of peer review at *Energy Economics*, gender gaps in reviewing and revision times decline. Assuming experience does not *reduce* one's ability to identify manuscript quality, our results suggest that if editors disproportionately assign female-authored papers to the most experienced referees, they may be able to increase equity in outcomes without reducing the objectivity or informativeness of the refereeing process. Moreover, monitoring compliance should be relatively straightforward since outcomes—in our case, referee allocation—are easily observable whereas biases, usually, are not (Lundberg 1991).

[1] We find authors spend slightly less time responding to reports by expert referees then they do responding to reports by novice referees (see Figure 3).

Third, we also contribute to an emerging body of empirical research studying discrimination in a dynamic setting. For example, Hengel (2019) uses the choices authors make over multiple submissions to study potential discrimination against female economists in academic peer review. Bohren *et al.* (2019) show that as evaluators are presented with more knowledge about posters' skills, they are less likely to discriminate against (and may even start discriminating in favour of) female posters in an online platform. In this paper, we observe a similar decline in gender gaps as evaluators' information increases, and like Bohren *et al.* (2019), we interpret our results as evidence of statistical discrimination. However our dynamics involve changes in evaluators' knowledge of and confidence in the process of peer review holding the information they have about the subjects of their assessments fixed. Thus our results point to an alternative—and probably more feasible—means of mitigating the negative consequences of bias.

Fourth, this paper builds on research suggesting women are subject to tougher standards than men (see *e.g.*, Card *et al.* 2020; Hengel 2019; Hengel and Moon 2020). Most relevant to our work, Hengel (2019) shows that female-authored papers published in two top general interest economics journals—*Econometrica* and the *Review of Economic Studies*—spend three to six months longer under review compared to observably equivalent male-authored papers. An important contribution of our study is to replicate the direction of this finding using more disaggregated data from a less prestigious field journal; our results therefore suggest that peer review may impose greater costs on female economists wherever they submit.

Finally, we also contribute to the broader literature studying editorial patterns (Card and DellaVigna 2013; Ellison 2002), bias in editorial decisions (Abrevaya and Hamermesh 2012; Bransch and Kvansnicka 2017; Card *et al.* 2020; Hospido and Sanz 2019) and female academics' lagging productivity and under-representation (Auriol *et al.* 2019; Bayer and Rouse 2016; Chari and Goldsmith-Pinkham 2017; Ductor *et al.* 2018; Gamage *et al.* 2020; Ginther and Kahn 2004; Heckman and Moktan 2019; Lundberg and Stearns 2019; Teele and Thelen 2017).

This paper proceeds in the following order. Section 2 describes the data. Section 3 builds the conceptual framework around which we interpret the results presented in Section 4. Section 5 concludes.

2 Data

Our dataset links administrative information stored in Elsevier's Editorial Manager (EEM) system with bibliographic and demographic data on the articles and authors published in *Energy Economics*, a peer-reviewed field journal focused on the areas of energy economics and energy finance. *Energy Economics* is published bi-monthly; in 2020, its Clarivate Analytics Journal Impact Factor and Scopus CiteScore were 7.042 and 10.0, respectively (*Energy Economics* 2021).

To proxy for a manuscript's "gender", we use the gender of its corresponding author, whom editors, tenure committees and other researchers generally assume contributed the most to the paper (see *e.g.*, Bhandari *et al.* 2003; Bhandari *et al.* 2004; Bhandari *et al.* 2014; Duffy 2017; Mattsson *et al.* 2011; Wren *et al.* 2007). Genders were assigned manually using the following hierarchy of information: (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.

We constructed our dataset in three steps. First, we extracted the following from EEM: (i) manuscript id numbers, titles and corresponding authors for full-length, regular issue papers published in *Energy Economics*; and separately, (ii) the names and unique id numbers for every editor and referee who reviewed a paper that was eventually accepted for publication in *Energy Economics*. In the second step, we collected publicly available information on manuscripts (*e.g.*, Web of Science citation counts), identified referees and authors with multiple EEM accounts and manually assigned a gender to every corresponding author. For the final step, we deployed a Python program to extract the following information from EEM: review time metrics (*e.g.*, manuscript submission dates and the number of days referees took to return their reports), editor and referee recommendations (*i.e.*, "Accept", "Major Revision", "Minor Revision" and "Reject") and round. The program then merged these data with the data collected in steps 1 and 2 and anonymised referees', editors' and authors' identities.

Our final dataset includes anonymised information on each round of review for 2,359 full-length, regular issue manuscripts submitted via the EEM system and published in *Energy Economics* on or before June 2019.[2] For 2,017 of these papers, the corresponding author was male; on 342, she was female.

Graph (A) in Figure 1 plots the number of manuscripts in our data by submission and publication year. The number of manuscripts published in *Energy Economics* has steadily increased since April 2005 (the first month *Energy Economics* managed submissions through EEM) and January 2006 (the first month a paper submitted through EEM was published in *Energy Economics*). The percentage of female-authored papers, however, is fairly flat (Figure 1, Graph (B)): in 2006, 14.3 percent of published papers had a female corresponding author; in 2018, 14.5 percent did.

Figure 1's second row of graphs plots the average number of rounds manuscripts go through before being accepted. Graph (C) shows this figure over time; Graph (D) displays its distribution by author gender.[3] Graph (C) suggests that manuscripts submitted in 2016 went through the same number of rounds as manuscripts submitted in 2006. According to Graph (D), one or two rounds of review is more common among male-authored papers; three or more rounds is more common for female-authored papers. (The half-sample mode for female-authored papers is 3 rounds; for male-authored papers, it is only 2.) On average, female-authored manuscripts go through 2.7 rounds, whereas male-authored manuscripts go through 2.6.[4]

In the third row of Figure 1, we show data on the total number of days manuscripts spend with referees. Graph (E) plots this figure over submission and publication years: referees spent slightly more time reviewing papers in 2010 than they did in 2005, but since then their average number of days has been relatively flat. Graph (F) displays the distribution of time spent with referees by manuscript gender. It suggests that the amount of time referees take to review papers correlates with the gender of their corresponding authors: on average, male- and female-authored papers spend 220 and 237 days with referees, respectively.

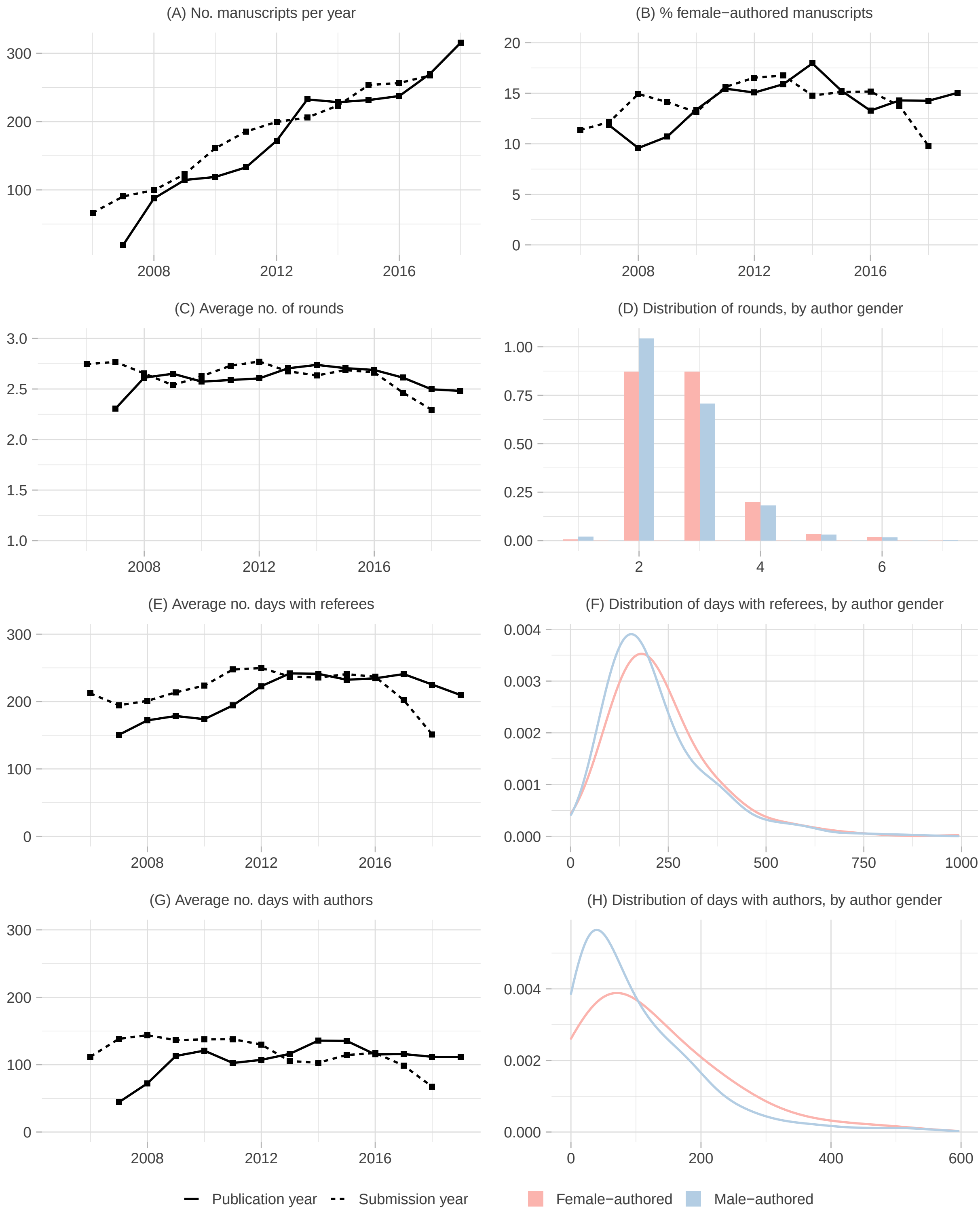
Female authors also spend more time revising than do male authors. Graph (H) plots the distribution of the total number of days authors take to revise their manuscripts by gender; on average, men spend 111 days revising their papers; women spend 140 days. According to Graph (G), revision time has not changed much in the last 10–15 years. In 2005, authors took, on average, 112 days to revise their papers; in 2016, they took 111 days.

In Appendix A, we summarise data on editorial decisions, proportion of female authors and average number of days spent with referees and authors in each round of review. Less than one percent of manuscripts are immediately accepted in the first round; three-quarters and one-quarter are asked to make major and minor revisions, respectively. Time spend reviewing and revising negatively correlates

[2] *Energy Economics* began using EEM in April 2005; we do not have data on manuscripts submitted before then. Additionally, for a very small number of manuscripts (fewer than 10), data are missing on a single round of review.

[3] Note that the slight decline and rise in average number of rounds (Graph (C))/time spent with referees (Graph (E))/time spent with authors (Graph (G)) when plotted over submission and publication year, respectively, reflect selection bias: manuscripts included in our data must have been submitted on or after April 2005 and published on or before June 2019. Thus, manuscripts that were published in 2006 or submitted in 2018 would have necessarily experienced a faster than average review process.

[4] A Kolmogorov-Smirnov test suggests that the two samples are not drawn from the same probability distribution ($D = 0.09$, p-value $= 0.01$).



Note. Graph (A) displays the two-year moving average of the total number of manuscripts published (solid line) and submitted (dashed line) each year. Graph (B) is the two-year moving average of the percentage of papers published and submitted each year that are female-authored. Graphs (C) and (D) plot the two-year moving average number of rounds per year and the distribution of rounds, by author gender, respectively. Graphs (E) and (F) plot the two-year moving average over time and distribution by gender of the total number of days a manuscript spends with referees, respectively. Graphs (G) and (H) show the two-year moving average over time and the distribution by gender of the total number of days a manuscript spends with authors.

Figure 1: Summary statistics on manuscript submissions to *Energy Economics*

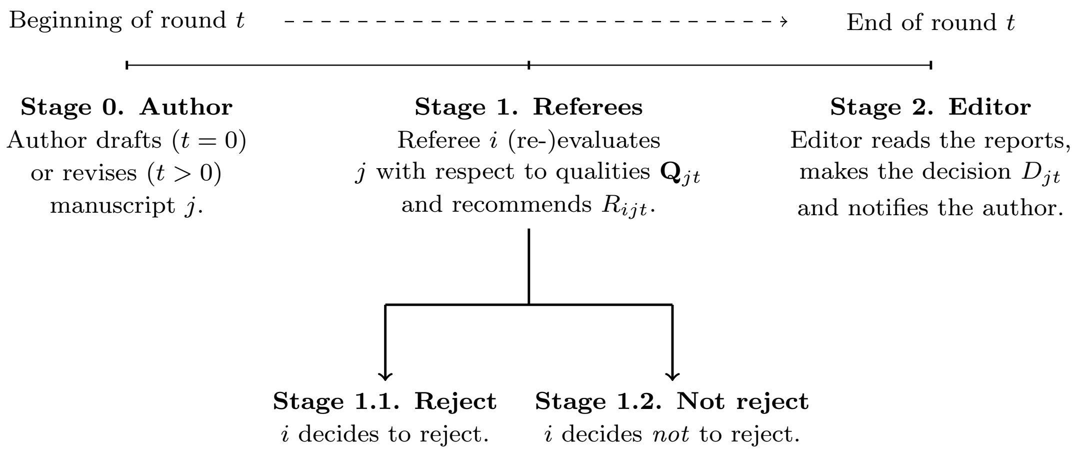


Figure 2: Sequence of events during revision round t

with round but female-authorship positively correlates with it. This later observation is consistent with evidence in Figure 1 suggesting that female-authored papers go through more rounds of review.

3 Conceptual framework

Figure 2 illustrates the sequence of events that occur during revision round t. Round t begins with the author either drafting ($t = 0$) or revising ($t > 0$) manuscript j and then (re-)submitting it to *Energy Economics*. When j is received by the editorial office, it is assigned to a handling editor who sends it to one or more referees, denoted by i. Each i evaluates a vector of j's qualities $\mathbf{Q}_{jt}$—*e.g.*, novelty, rigour and readability—and summarises his opinion of them in the report and recommendation R_{ijt}. t concludes with the editor reading and processing R_{ijt} for all i and making the round-specific decision D_{jt}. If D_{jt} is accept or reject, then the author is notified and peer review ends; otherwise, he is asked to revise j and round $t + 1$ begins.

In this paper, we focus on two nodes in Figure 2: (i) the time authors spend revising their papers after receiving a revise and resubmit decision (stage 0 for $t > 0$); and (ii) the time referees take to review papers that they do not reject (stage 1.2). The former decision is a practical one—we do not observe the time authors take to initially draft their papers. The latter decision was made because referees spend more time reviewing papers they *don't* reject than they spend reviewing papers they *do* reject, suggesting a non-monotonic relationship between R_{ijt} and evaluation time. (See Appendix B for evidence.) We hope to explore gender differences at stages 1.1 and 2 in future work.

3.1 Referees

Consider first the difference in the amount of time referees take to evaluate papers by women versus papers by men, conditional on non-rejection (stage 1.2, Figure 2):

$$\Delta^R = \mathbb{E}\Big[\text{time}^R_{ij_f t} - \text{time}^R_{ij_m t} \Big| R_{ij_l t} \neq \text{Reject}, \mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}\Big], \tag{1}$$

where $\text{time}^R_{ij_l t}$ is the length of time referee i spends evaluating manuscript j_l in round t and $l \in \{m, f\}$ denotes female- and male-authored papers, respectively. Δ^R is restricted to referee reports that did not recommend rejection ($R_{ij_l t} \neq$ Reject); because i's decision to request a revision may itself be discriminatory, it also conditions on the criteria that determines that selection (*i.e.*, $\mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}$).[5]

$\Delta^R > 0$ could be caused by several factors, most of which can be framed as some form of discrimination against female authors. The first, traditional taste-based discrimination, includes behaviour triggered by prejudicial preferences, *e.g.*, a stronger desire to retaliate against female peers (see, *e.g.*, Dehdari *et al.* 2019; Rehg *et al.* 2008) or greater reluctance—possibly motivated by envy—to advance their careers (for related evidence, see, *e.g.*, Ratliff and Oishi 2013). It also covers attitudes about gender that cause the people who hold them to worry less about the negative externalities their actions impose on women—*e.g.*, because referees respect women's time less or aren't as concerned about the impact a delayed report will have on their careers.

Imperfect information drives the second form of discrimination: statistical discrimination. Referees may spend more time evaluating female-authored manuscripts because they believe—correctly or incorrectly—that women submit lower quality work (Phelps 1972); alternatively, for a variety of reasons they could find women's writing style or research methods harder to interpret (Aigner and Cain 1977). Statistical discrimination from referees might also be prompted by statistical—or even taste-based—discrimination from other actors. For example, referees may scrutinise women's papers more closely because they believe that editors or readers are more critical of their work.

A third type of discrimination relates to power. Most organisations have developed over time to accommodate predominantly male decision-makers. As a result, existing institutional norms and practices can often subtly (and sometimes not-so-subtly) disadvantage women.[6] For example, referees could rely on seminar and conference presentations to learn about the papers they review, but childcare responsibilities may mean women are less able to attend these events. Alternatively, refereeing norms may have evolved to make university affiliation an acceptable proxy for manuscript quality; however, hiring discrimination, family commitments or other gender-specific constraints probably mean that affiliation does not convey the same signal about the quality of a woman's paper that it does about the quality of a man's.

Moreover, each form of discrimination overlaps with and feeds back into other forms of discrimination. Long-standing taste-based and statistical discrimination are undoubtedly codified into institutional practices and even the law (*e.g.*, red-lining in U.S. real estate markets (Hillier 2003; Hillier 2005)). Both also stymie women's ability to influence their professions, thus potentially setting in motion further discrimination, *e.g.*, by reducing the relative rewards from mentoring women, facilitating their careers or even writing them thoughtful, timely reports. Furthermore, widespread gender disparities probably spawn inaccurate stereotypes (Lang and Spitzer 2020; Reskin 2012); for example, correlation between gender and organisational rank combined with the latter's use as a proxy for quality may, as a result, cause some referees to mistakenly infer that gender *also* proxies for quality.

3.2 Authors

Now consider gender differences in the amount of time authors spend revising their papers (Figure 2, stage 0 for $t > 0$):

$$\Delta^A = \mathbb{E}\Big[\text{time}^A_{j_f t} - \text{time}^A_{j_m t} \,\big|\, D_{j_l t-1} \notin \{\text{Accept}, \text{Reject}\}, \mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}\Big], \tag{2}$$

where $\text{time}^A_{j_l t}$ is the length of time spent revising manuscript j_l in rounds $t > 0$ and $l \in \{f, m\}$ denotes the gender of j's corresponding author (male and female, respectively). Δ^A conditions on $D_{j_l t-1} \notin \{\text{Accept}, \text{Reject}\}$ to account for the fact that time^A_{jt} is only observed if revisions were requested in the

[5] By conditioning on $\mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}$, we additionally account for selection issues introduced when manuscripts were first submitted to *Energy Economics*—*e.g.*, because female authors submit higher quality papers.

[6] Often, these practices and norms were neither driven by prejudice nor were the managers, directors, or employees following the norms themselves gender prejudiced (for a discussion, see Small and Pager 2020).

previous round; $\mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}$ adjusts for the resulting selection bias.

Both discriminatory and non-discriminatory factors contribute to $\Delta^A > 0$. The first group includes forms of discrimination that are independent of information—*i.e.* pure taste-based and institutional discrimination. For example, prejudicial preferences can mean referees respect women's time less and/or aren't as concerned about the negative externalities low quality reports impose on female authors; alternatively, women are less likely to become, *e.g.*, journal editors, so referees may not have enough of an incentive to write them high quality reports.

The second group of factors relates to gender differences in the amount of time authors have available to revise their papers. Greater childcare obligations, teaching loads or service responsibilities may force women to push their eventual resubmission dates further into the future. Additionally, fewer grants mean women will be less able to hire project or research assistants, leaving them with fewer resources to complete a revision. Evidence also suggests that female academic economists are more precariously employed (see *e.g.*, Bateman *et al.* 2021); as a result they may have to spend more time searching for and applying to jobs.

The third set of factors is caused by incomplete information on the part of referees. Referees may ask for more changes to female-authored papers because they believe—again, correctly or incorrectly—that their quality is lower (statistical discrimination). Alternatively, referees may receive a signal about the quality of female-authored papers that is less precise—perhaps because they find them harder to interpret or further away from their own expertise (institutional or statistical discrimination). Since less knowledgeable individuals appear to generate vaguer, less detailed explanations when asked to justify their solutions to a given task (see *e.g.*, Means and Voss 1996), women may therefore find themselves responding to referee reports that are, on average, more difficult to decipher and challenging to implement.

A final group of factors also revolves around incomplete information, but this time on the part of authors. Women may be more poorly informed about the appropriate way to interpret referee reports—perhaps because male advisers are less willing to convey that information to their female students or their female students are less likely to ask them for it.[7] Alternatively, higher risk aversion, greater conscientiousness or a larger marginal cost of rejection may mean women work harder (and take more time) revising their papers, conditional on receiving the same (incomplete) information as a man.

3.3 The role of referee experience

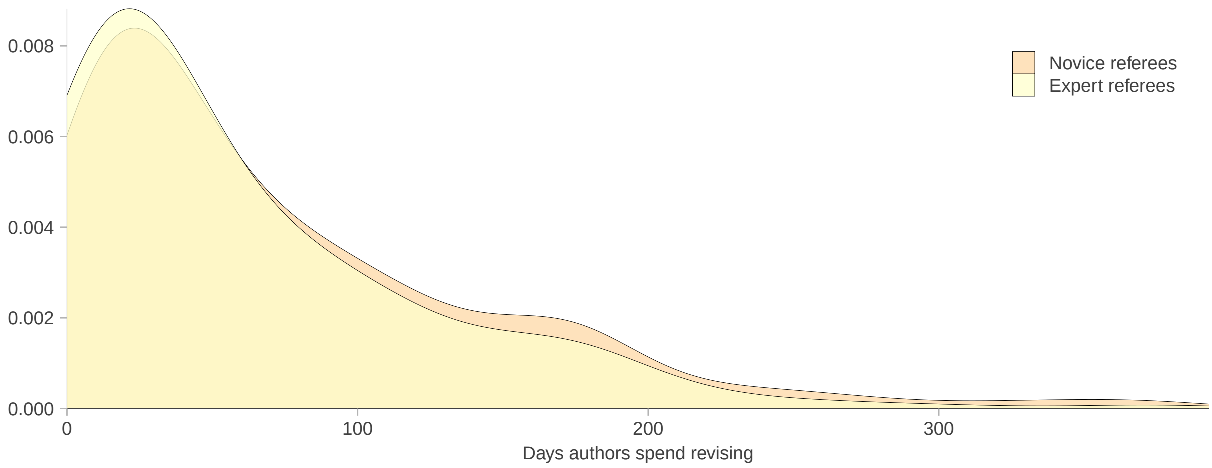
A key factor driving gender gaps in reviewing and revision times involves incomplete information—especially on the part of referees. More specifically, if referees had perfect information, then they could not underestimate the quality of female-authored papers nor could the signals they receive about them be any less precise than the signals they receive about male-authored papers. Thus, in this world $\Delta^R > 0$ indicates pure taste-based or institutional discrimination.

Furthermore, better informed referees probably trickle down to better informed authors. That is, perfectly informed referees should be able to clearly specify the changes that would need to be made before a paper could be published. As a result, authors should have no trouble interpreting their reports; $\Delta^A > 0$ would therefore probably be caused by differences in the time and resources men and women have available to revise their papers.[8]

Although information is neither complete nor costless to obtain, experienced referees are probably better at distinguishing between low and high quality papers compared to inexperienced referees. They are also

[7] Both scenarios can be view as forms of institutional discrimination, since information is conveyed in a way that systematically disadvantages women.

[8] In the presence of complete information, $\Delta^A > 0$ could also result from taste-based discrimination on the part of referees; however, we find it implausible that a perfectly informed gender-prejudiced referee would prefer intentionally writing a vague and difficult-to-implement report instead of simply rejecting the paper.



Note. Figure plots the distribution of time spent revising in the first round of review for papers reviewed by novice (orange) and expert (yellow) referees (defined as having reviewed fewer than five and at least five papers for *Energy Economics*, respectively). Sample restricted to manuscripts for which time spent revising did not exceed 400 days.

Figure 3: Distribution of time spent with author by referee experience

undoubtedly more knowledgeable about the standards of acceptance at *Energy Economics*—indeed, by frequently refereeing they regularly set those standards.[9] Consequently, the signal an experienced referee receives about the quality of a submitted paper is probably more accurate—and contains fewer gender-specific distortions—than the signal received by a novice referee. Furthermore, signal clarity arguably gives rise to more precise and easier-to-implement reports, all else equal.

Figure 3 supports this hypothesis. On average, authors spend slightly less time responding to reports by expert referees than they do responding to reports by novice referees, where "novice" and "expert" are defined as having reviewed four or fewer and five or more eventually accepted papers for *Energy Economics*, respectively.

We use this insight to identify the impact of referee information on gender gaps in reviewing and revision time. Let $s \in \{A, R\}$ denote stages 0 (with authors) and 1.2 (with referees) of review, n and e novice and experienced referees, respectively, and define $\tilde{\Delta}^s$ as follows

$$\tilde{\Delta}^s = \Delta^s_e - \Delta^s_n, \tag{3}$$

where Δ^R_e is Equation (1) for experienced referees, Δ^A_e is Equation (2) for papers reviewed by experienced referees, *etc.*

To understand how Equation (3) separates factors related to incomplete information from those that are independent of it, suppose $\Delta^R_n > 0$ and $\tilde{\Delta}^R < 0$. This pattern suggests novice referees initially statistically discriminate against female authors; with experience, however, their ability to identify quality in men's and women's papers converges, so Δ^R_e is smaller than Δ^R_n. On the other hand, $\tilde{\Delta}^R \geq 0$ suggests that referees spend longer with female-authored papers regardless of experience, which is consistent with pure taste-based or institutional discrimination. Similarly, $\Delta^A_n > 0$ and $\tilde{\Delta}^A \geq 0$ indicates no negative relationship between Δ^A and referee experience; instead, female authors probably have less time and fewer resources available to revise their papers. Conversely, $\tilde{\Delta}^A < 0$ suggests that Δ^A declines as referee experience increases. This is likely because inexperienced referees write reports that female authors have an especially hard time responding to—*e.g.*, because women are more risk averse or inexperienced

[9] According to Berk *et al.* (2015, p. 2), "The job of the referee is to provide expert and unambiguous advice to the editor about whether or not a paper is publishable." Unfortunately, the meaning of "publishable" depends on how other referees, editors and readers of *Energy Economics* would categorise a paper in relation to others it has (or likely will) publish. (See Lipman (2009) for a related discussion.)

referees write them particularly tough reports (statistical discrimination).

4 Empirical strategy and results

4.1 Referees

4.1.1 Estimation strategy

To estimate Δ^R, we would ideally regress the amount of time referees spend evaluating papers on author gender and manuscript quality:

$$\text{time}^R_{ijt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \,t + \tau_j + \varepsilon_{ijt}, \tag{4}$$

where time^R_{ijt} is the length of time (in days) i spends evaluating manuscript j in round t, female_j is a binary variable equal to one if j's submitting author is a woman, $\mathbf{Q}_{jt}$ is a hypothetical vector of attributes that perfectly capture j's quality and ε_{ijt} is the error term.

Applying Equation (4) to the data poses two problems. First, β_1 consistently estimates Δ^R only if referee assignment in round t does not depend on author gender conditional on $\mathbf{Q}_{jt}$. For this reason, we additionally control for year of submission (τ_j) and t: Figure 1 suggests time^R_{ijt} and female_j have possibly changed over the years; female-authored papers are also more likely to go through multiple rounds of review, and time^R_{ijt} is shorter in later rounds.[10]

Our second problem is that we do not know $\mathbf{Q}_{jt}$. We proxy with three (imperfect) measures instead—citations, R_{ijt} and D_{jt}. Because citations are neither round-specific nor measured pre-treatment, they may be influenced by peer review in a way that correlates with female_j. For example, referees could work harder to increase the quality of women's manuscripts—*i.e.*, conditional on citations, $\mathbf{Q}_{j_f t} < \mathbf{Q}_{j_m t}$—so β_1 over-estimates the true value Δ^R. On the other hand, R_{ijt} and D_{jt} are probably affected by the same discrimination captured by Equation (1)—*e.g.*, $R_{ij_f t} < R_{ij_m t}$ conditional on $\mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}$—in which case β_1 likely *under*-estimates Δ^R.

Despite these limitations, it is still possible to conservatively estimate the true value of Δ^R if at least one proxy is not biased *in favour* of women, conditional on $\mathbf{Q}_{j_f t} = \mathbf{Q}_{j_m t}$. According to past research, this assumption probably holds. Men are disproportionately more likely to cite their own (King *et al.* 2017) and other male-authored work (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Koffi 2019). Economists believe female-authored papers are cited less, holding quality constant (Card *et al.* 2020). Card *et al.* (2020) also find that $R_{ij_f t} < R_{ij_m t}$ and $D_{j_f t} < D_{j_m t}$, conditional on citations.

As we emphasise in Section 3.1, each form of discrimination probably reinforces others and the boundaries dividing them are anyway fuzzy. For these reasons, Equation (4) does not rigorously isolate the specific types of discrimination potentially driving $\Delta^R \neq 0$; in order to better inform this debate, we also estimate Equations (5) and (6). Equation (5) investigates how the gender gap in time spent with referees changes with observable controls:

$$\text{time}^R_{ijt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \,t + \tau_j + \mathbf{X}_{jt} + \varepsilon_{ijt}, \tag{5}$$

where $\mathbf{X}_{jt}$ is a vector of editor, referee, field and institution fixed effects. Equation (6) is the practical implementation of Equation (3):

$$\text{time}^R_{ijt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \,t + +\beta_4 \,\text{exp}_{\cdot ij} + \beta_5 \,\text{female}_j \times \text{exp}_{\cdot ij} + \tau_j + \varepsilon_{ijt}, \tag{6}$$

[10] See Figure 1 for evidence suggesting female authors go through more rounds of review. The coefficient on t (round) from a regression of time^R_{ijt} on it suggests referees take on average 15.9 fewer days each additional round of review (standard error 0.99).

where exp._{it} is the number of accepted papers referee i previously reviewed for *Energy Economics*.

Assume referee assignment across author gender is orthogonal to the former's experience, referees are better at identifying "publishable" manuscripts as they review and accept more papers for *Energy Economics* and our proxy for $\mathbf{Q}_{jt}$ is not biased in favour of women. Then the combination of a positive β_1 and non-negative β_5 in Equation (6) suggests that Δ^R does not decline as referees gain experience; it is therefore likely due to taste-based discrimination. A positive β_1 together with a negative β_5, however, suggests that statistical discrimination on the part of inexperienced referees at least partly explains why they spend longer evaluating female-authored papers.

4.1.2 Results

The first panel of Table 1 displays results from OLS estimation of Equation (4). Column (1) proxies for $\mathbf{Q}_{jt}$ using citations alone; it suggests that referees take 4.4 days longer to review female-authored papers. In columns (2)–(4), we control for R_{ijt}, D_{jt} and all three proxies together; results are similar to those in column (1). Coefficients on the remaining co-variates correspond to intuition: more highly cited papers are reviewed slightly faster as are papers being reviewed in later rounds. Referees are also quicker to accept than they are to recommend a revision. (The coefficients on "revise (major)" and "revise (minor)" are relative to the base level "accept".)

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 "Oster bounds" corresponding to the assumption that the unobservables explain about as much of the variation in time^R_{ijt} as the observables do.[11] They suggest referees spend 4–5 days longer reviewing women's papers.

Columns (5)–(7) re-estimate Equation (5) controlling for editor, referee and field fixed effects. β_1 resembles those reported in columns (1)–(4); it is also robust to controlling for author prominence, number of co-authors and manuscript length (see Appendix E).

Column (8) controls for institution fixed effects. It is interesting for two reasons: first, β_1 doubles; second, institution fixed effects absorb substantial variation in time^R_{ijt}—the R^2 jumps 20 points between columns (7) and (8). These results suggest an underlying association between institutions and paper-specific unobservables that could partially bias our estimate of β_1 in Equation (4)—*e.g.*, referees could be faster to review papers by people they personally know, and they may be more likely to know female authors conditional on non-rejection. Indeed, when we re-estimate Equation (4) with institution fixed effects, the coefficient on female is 6.3 (standard error 2.4); its corresponding Oster bounds suggest referees take 6–8 days longer to review women's papers.

To test whether gender differences in time^R_{ijt} are related to institutional rank, we re-estimate Equation (5) controlling for the prestige of an author's affiliation and its interaction with female_j.[12] Figure 4 shows our results: as prestige increases, referees take less time to review male-authored papers; for women, however, time^R_{ijt} and rank are positively related. In other words, referees are quicker to review male-authored papers—but *slower* to review female-authored papers—from higher ranked universities.

The final column of Table 1 displays results from OLS estimation of Equation (6). It reports a positive β_1 and negative β_5. Assuming referee assignment across author gender is orthogonal to the former's experience, referees are quicker at identifying quality as they review more papers for *Energy Economics*

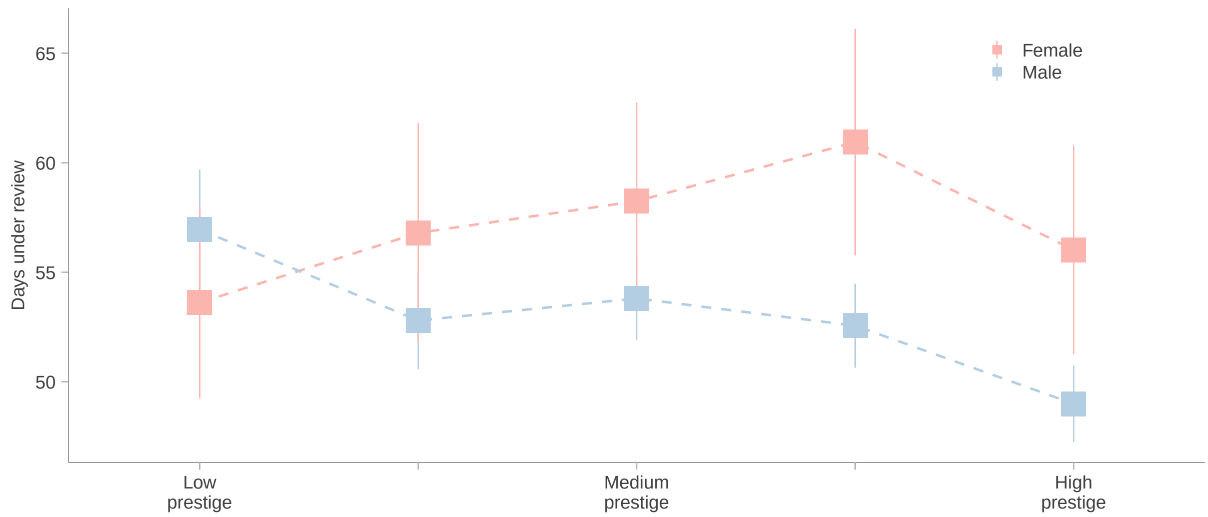
[11] Specifically, we assume that the R^2 from a regression of the observables *and* unobservables on time^R_{ijt} is no more than $2\,\widetilde{R}^2 - \mathring{R}^2$, where $\widetilde{R}^2$ is the R^2 from estimating Equation (4) (*i.e.*, a regression of the observables on time^R_{ijt}) and $\mathring{R}^2$ is the R^2 from only regressing female_j on time_{ijt}.

[12] Since we are interested in the prestige of an institution in the field of energy economics, ranking is determined by the number of manuscripts published in *Energy Economics* with a corresponding author affiliated with the institution. We then grouped institutions into five roughly equally sized groups.

Table 1: Gender gap in the number of days spent with referees

	Eq. 4				Eq. 5				Eq. 6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female (β_1)	4.417***	4.753***	4.725***	4.095**	4.449***	4.169***	3.308**	8.282***	9.768***
	(1.658)	(1.696)	(1.692)	(1.638)	(1.601)	(1.594)	(1.665)	(2.636)	(2.903)
t (round)	−15.896***	−12.930***	−13.553***	−12.704***	−15.050***	−13.332***	−13.224***	−13.516***	−13.557***
	(0.973)	(1.346)	(1.632)	(1.57)	(0.908)	(1.278)	(1.252)	(1.201)	(1.203)
citations (asinh)	−5.244***			−5.216***	−4.699***	−2.973***	−3.128***	−2.919***	−3.031***
	(0.837)			(0.832)	(0.846)	(0.988)	(0.868)	(0.901)	(0.895)
R_{ijt} (referee's recommendation)									
revise (major)		8.502***		6.781***					
		(1.956)		(2.238)					
revise (minor)		6.568***		5.216***					
		(1.515)		(1.762)					
D_{it} (editor's decision)									
revise (major)			8.099***	3.230					
			(2.664)	(3.078)					
revise (minor)			5.738***	3.114					
			(1.698)	(1.943)					
Referee experience									
experience									0.216**
									(0.108)
experience×female									−0.107**
									(0.05)
No. obs.	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035
R^2	0.083	0.070	0.069	0.087	0.089	0.093	0.114	0.313	0.315
Oster bounds (β_1)	[4.2, 4.4]	[4.8, 4.9]	[4.7, 4.8]	[3.5, 4.1]	–	–	–	–	–
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
Referee						✓	✓	✓	✓
JEL (secondary)							✓	✓	✓
Institution								✓	✓

Note. Figures correspond to coefficients from estimating Equations (4), (5) and (6). The dependent variable is the amount of time the referees took to evaluate a paper. Standard errors clustered by referee in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.



Note. Figures represent time spent with referees for male (blue) and female (pink) authors at low- to high-ranked institutions. Results are from estimating Equation (5), controlling for citations (asinh), institutional rank and the latter's interaction with gender. See Footnote 12 for details on how the categorical variable of institutional prestige was constructed. Lines represent 90 percent confidence intervals; robust standard errors in parentheses.

Figure 4: Gender differences in time_{ijt}^{R} across authors' institutional prestige

and citations are not biased in favour of women, conditional on $\mathbf{Q}_{jt}$, this suggests that gender differences in the time referees take to evaluate papers declines as they gain experience.[13] As discussed in Section 3.3, this pattern of behaviour is consistent with statistical discrimination on the part of inexperienced referees.

In Figure 5 we re-estimate Equation (6) but replace the continuous measure of referee experience with a categorical one that accounts for the following types of referees: novices (defined as having refereed four or fewer papers (including the current paper) that were eventually published in *Energy Economics*), experts (5–19 papers) and so-called "power referees" (20 or more papers).[14] The gender gap is about equally large among novices and experts but drops for power referees. These results hold after accounting for referee fixed effects, indicating that individuals become more accurate at evaluating papers as they gain experience reviewing for *Energy Economics*. Nevertheless, Figure 5 suggests that referees may require a lot of experience before Δ^R declines.

4.2 Authors

4.2.1 Estimation strategy

To estimate Δ^A, we regress the amount of time authors take to complete their revisions on their gender and the quality of their manuscripts:

$$\text{time}_{jt}^{A} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \,t + \tau_j + \varepsilon_{jt}, \tag{7}$$

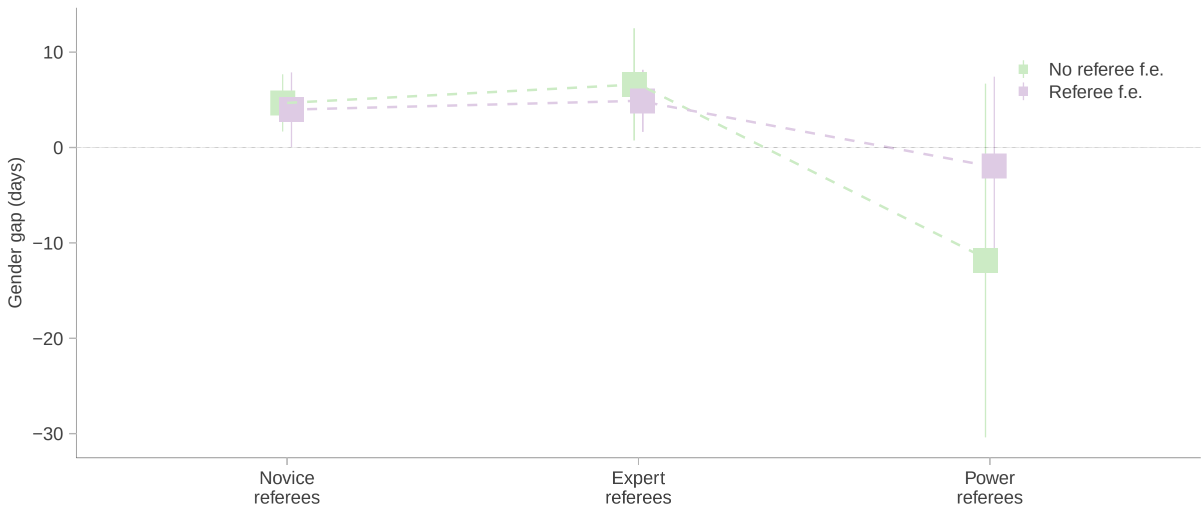
where time_{jt}^{A} is the number of days between the date paper j's corresponding author is notified of the editor's decision and the date she re-submits her paper. As in Equation (4), we control for year of submission and t; to proxy for $\mathbf{Q}_{jt}$, we use D_{jt-1}, D_{jt} and citations (asinh).[15]

A positive Δ^A suggests women spend longer revising their papers but does not tell us why (see Section

[13] See Appendix C for evidence that referee assignment is independent of author gender conditional on experience.

[14] While their numbers are small, power referees review a non-negligible portion of papers submitted to *Energy Economics*.

[15] Given Equation (7) is estimated at the author-level, we do not proxy for $\mathbf{Q}_{jt}$ using R_{itj}, which is referee-specific. Nevertheless, disaggregating the data to the referee level or controlling for referees' average scores produce similar results (available on request).



Note. Graph displays the coefficients on female from a re-estimation of Equation (6) that accounts for experience using the following three categories: "novice" referees (referees who have refereed four or fewer papers (including the current paper) for *Energy Economics*), "experienced" referees (5–19 papers) and "power" referees (20 or more papers). Lines represent 90 percent confidence intervals; standard errors clustered at the referee level.

Figure 5: The relationship between referee experience and Δ^R

3.2). To investigate, we additionally estimate Equations (8) and (9). Equation (8) considers how β_1 changes with observable controls:

$$\text{time}_{jt}^{A} = \beta_0 + \beta_1\,\text{female}_j + \beta_2\,\mathbf{Q}_{jt} + \beta_3\,t + \tau_j + \mathbf{X}_{jt} + \varepsilon_{jt}, \tag{8}$$

where $\mathbf{X}_{jt}$ is a vector of editor, referee, field and institution fixed effects. Equation (9) uses referee-level data to estimate $\tilde{\Delta}^A$ (Equation (3)):

$$\text{time}_{ijt}^{A} = \beta_0 + \beta_1\,\text{female}_j + \beta_2\,\mathbf{Q}_{jt} + \beta_3\,t + \beta_4\,\text{exp.}_{ij} + \beta_5\,\text{female}_j \times \text{exp.}_{ij} + \tau_j + \varepsilon_{ijt}, \tag{9}$$

where time_{ijt}^{A} is the amount of time paper j's authors take in round t to respond to referee i's report from round $t-1$ and exp._{ij} is the number of accepted papers i previously reviewed for *Energy Economics*.

To interpret Equation (9), suppose the following: (i) referee assignment across author gender is orthogonal to the former's experience; (ii) our proxy for $\mathbf{Q}_{jt}$ is not biased in favour of women; and (iii) more experienced referees write reports that authors find easier to implement. When these assumptions are satisfied, then the combination of a positive β_1 and negative β_5 suggests that the gap between men's and women's revision times declines when reviewed by experienced referees. As argued in Section 3.3, this is likely because inexperienced referees write reports that female authors find especially difficult to respond to—either thanks to a combination of unclear requests and women's greater risk aversion/lower confidence/*etc.* or because inexperienced referees statistically discriminate against female authors and, as a consequence, write them tougher reports. On the other hand, a positive β_1 and non-negative β_5 suggests that Δ^A does *not* decline as referees' information gets better. This points instead to women having less time and fewer resources available to revise their papers or referees having less of an incentive (or desire) to write them high quality reports (pure taste-based or institutional discrimination).

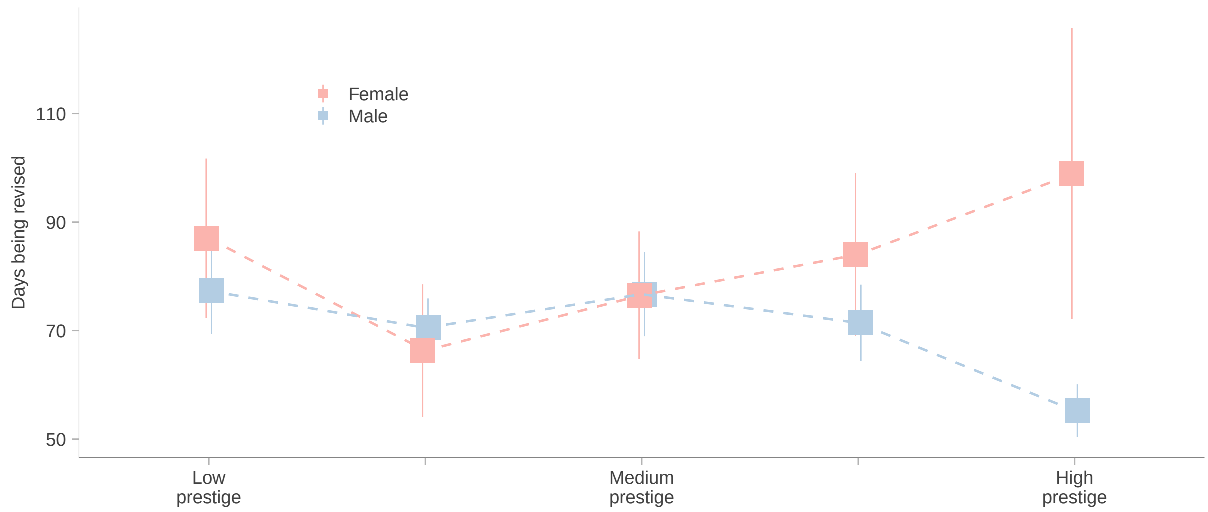
4.2.2 Results

The first panel of Table 2 displays results from OLS estimation of Equation (7). Column (1) proxies for $\mathbf{Q}_{jt}$ using citations. It suggests that in each round of review, female authors take 12.3 days longer revising their papers. Results are similar when controlling for D_{it-1}, D_{it} and all three quality proxies

Table 2: Gender gap in the number of days spent with authors

	Eq. 7				Eq. 8				Eq. 9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	12.282**	12.467**	12.900**	10.899**	11.514**	11.430**	23.575***	19.522*	24.660**
	(5.316)	(5.515)	(5.569)	(5.295)	(5.352)	(5.277)	(8.032)	(10.734)	(10.837)
t	−41.537***	−28.216***	−38.156***	−27.842***	−42.336***	−43.428***	−45.079***	−58.873***	−58.823***
	(2.372)	(2.279)	(2.37)	(2.204)	(2.383)	(2.439)	(3.045)	(3.212)	(3.223)
citations (asinh)	−12.444***			−12.379***	−13.182***	−12.524***	−12.620***	−12.563***	−12.459***
	(1.842)			(1.765)	(1.857)	(1.845)	(2.793)	(3.206)	(3.192)
D_{it-1} (revise (major))		46.993***		44.464***					
		(3.674)		(4.206)					
D_{it}									
revise (major)			25.797***	3.232					
			(7.211)	(7.409)					
revise (minor)			22.514***	5.905					
			(3.982)	(4.478)					
Referee experience									
experience									−0.043
									(0.163)
experience×female									−0.402**
									(0.19)
No. obs.	3,814	3,814	3,809	3,809	3,814	3,814	3,814	6,440	6,440
R^2	0.112	0.133	0.105	0.151	0.114	0.153	0.394	0.443	0.444
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
JEL (secondary)						✓	✓	✓	✓
Institution							✓	✓	✓
Referee								✓	✓

Note. Figures correspond to coefficients from estimating Equations (7), (8) and (9). The dependent variable is the amount of time authors spent revising their papers. Standard errors clustered at the author level in columns (1)–(7) and at the referee level in columns (8)–(9). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.



Note. Gender differences in the time spent revising for male- (blue) and female-authored (pink) manuscripts at low- to high-ranked institutions. Results from estimating Equation (8), controlling for citations (asinh), institutional rank and the latter's interaction with gender. See Footnote 12 for details on how the categorical variable of institutional prestige was constructed. Lines represent 90 percent confidence intervals; robust standard errors in parentheses.

Figure 6: Gender differences in time_{jt}^{A} across authors' institutional prestige

together (columns (2)–(4)). Coefficients on these and other co-variates suggest that authors spend less time revising more highly cited papers and more time revising in earlier rounds. They also take longer when editors ask for major (instead of the baseline minor) changes.[16]

In Appendix E we show that the existence and magnitude of the gender gap in revision time is robust to controlling for corresponding author prominence, the prominence of the most prominent co-author, number of co-authors and manuscript length. In Table 2's columns (5) and (6), we also control for editor and field fixed effects. In all cases, β_1 resembles estimates in columns (1)–(4), tentatively suggesting that none drive women's longer revision times.

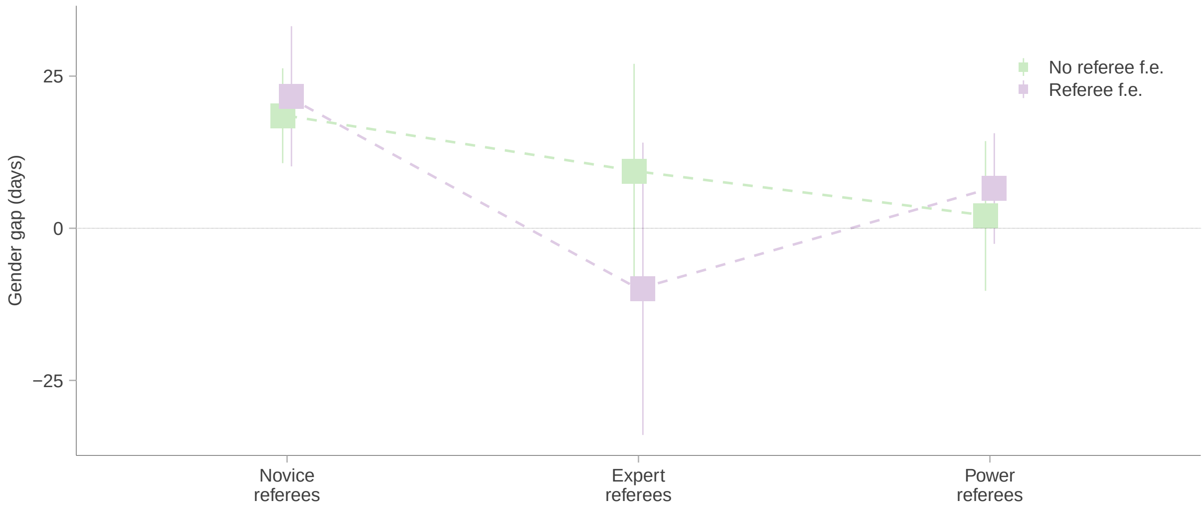
As in Section 4.1.2, adding institution fixed effects leads to a noticeable jump in β_1 and the regression's R^2 (column (7)). Moreover, gender differences in time_{jt}^{A} again positively correlate with an institution's rank. Figure 6 shows the amount of time men and women spend revising their papers over institutional prestige: there is no considerable gender difference among authors with less prestigious affiliations; for those from more prestigious universities, however, time_{jt}^{A} is much higher for women than it is for men.

Columns (8) and (9) of Table 2 use referee-level data.[17] Column (8) re-estimates Equation (8) including referee fixed effects; the coefficient on female_j roughly resembles its estimate in column (7). Column (9) displays results from OLS estimation of Equation (9). β_1 is positive and β_5 is negative, suggesting that Δ^A declines when papers are reviewed by experienced referees. As argued in Section 3.3, this is likely because inexperienced referees write reports that female authors find especially difficult to respond to, either because their reports for women are simply tougher to implement (statistical discrimination) or because women are more careful about responding to uncertain, vague or unclear requests from referees (*e.g.*, because they're more risk averse).

Figure 7 re-estimates Equation (9) but replaces the continuous measure of referee experience with the same categorical one used in Figure 5. The gender gap in revision time is largest when authors are reviewed by novice referees and smallest when reviewed by expert and power referees. In fact, after including referee fixed effects, it appears to be close to zero already among expert referees. These results

[16]Since time_{jt}^{A} is only observed if revisions were requested in the previous round, D_{jt-1} is never "accept".

[17]The referee-level data used in Table 2 is about 8 percent smaller than the sample used to estimate results in Table 1 because we do not observe the dependent variable for papers that were accepted after the first round.



Note. Graph displays the coefficients on female from a re-estimation of Equation (9) that accounts for experience using the following three categories: "novice" referees (referees who have refereed four or fewer papers (including the current paper) for *Energy Economics*), "experienced" referees (5–19 papers) and "power" referees (20 or more papers). Lines represent 90 percent confidence internals; standard errors clustered at the referee level.

Figure 7: The relationship between referee experience and Δ^A

indicate that individuals become better at writing reports—in the sense that female authors spend less time responding to them—as they gain experience refereeing for *Energy Economics*. They also suggests that it may be possible to reduce or eliminate Δ^A by allocating female-authored papers to referees with only a modest amount of experience.

5 Conclusion

This paper describes and analyses gender differences in the amount of time academics spend in peer review. Using detailed administrative data from an economics field journal, we find that in each round of review, referees spend 4.4 more days reviewing female-authored papers and female authors spend 12.3 more days revising their manuscripts. Both gaps remain constant or increase after including fixed effects for editors, referees, secondary *JEL* codes and authors' institutions. They are also robust to controlling for author prominence, manuscript length and number of co-authors.

A back-of-the-envelope calculation based on these results suggests that accepted female-authored papers take 3–4 weeks longer in review compared to similar quality male-authored papers.[18] Or in other words, for every dozen papers a woman publishes in journals like *Energy Economics*, she will have spent almost a year longer under review compared to a man with similar research quality and quantity. And given evidence that the total time gap is much larger at more prestigious journals (Hengel 2019), the cumulative gendered cost of peer review may be even higher.

But our evidence also suggests that journals may be able to ease this burden by disproportionately assigning female-authored papers to the most experienced referees. In particular, we find that both gender gaps are greater when papers are reviewed by novice referees than when they are reviewed by expert referees, conditional on referee fixed effects. Indeed, among highly experienced referees, the gaps either disappear or reverse. We argue that these patterns suggest novice referees initially statistically discriminate against female authors, but as they gain experience, their ability to evaluate men's and women's papers converges, so the gaps in reviewing and revising times decline.

[18] Calculation assumes 1.7 rounds of review (see Section 2). In Appendix D, Table D.3, we estimate the total gender gap; it suggests female authors spend 5–6 weeks longer in peer review compared to similar quality male-authored papers.

Unfortunately, our data cannot precisely identify *why* novice referees statistically discriminate. The evidence presented in Section 4 suggests that it is not directly connected to field or institution. It also appears to be unrelated to author prominence (Appendix E.1). But there are many remaining possibilities: for example, novice referees may have biased beliefs about female authors; alternatively, smaller networks may mean referees are simply less familiar with women's work (see *e.g.*, Ductor *et al.* 2018).

Finally, we believe our results suggest a need to expand the pool of competent and experienced referees. Conscientious editors may be reluctant to assign more manuscripts to expert referees for fear of over-stretching their most productive reviewers. Publishers can support them by training new referees. A notable example is the Institute of Physics, which offers a peer review certification programme to potential referees to help them gain confidence reviewing for its journals.

We are optimistic that our proposed policy levers can ease the burden statistical discrimination has on female authors without reducing the objectivity or informativeness of the refereeing process. A more diverse network of experienced reviewers may also improve the quality and relevance of economic research. Nevertheless, no single policy agenda is ever likely to completely fix what is obviously a complex problem. We therefore conclude by encouraging journals to gather further evidence on and conduct more rigorous evaluation of the gendered cost of peer review.

References

Abrevaya, J. and D. S. Hamermesh (2012). "Charity and favoritism in the field: are female economists nicer (to each other)?" *Review of Economics and Statistics* 94 (1), pp. 202–207.

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

Aldercotte, A. *et al.* (2017). *ASSET 2016: experiences of gender equality in STEMM academia and their intersections with ethnicity, sexual orientation, disability and age*. Report on the Athena Survey of Science, Engenieering and Technology 2017.

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.

Auriol, E., G. Friebel, and S. Wilhelm (2019). "Women in European economics". Mimeo.

Bateman, V., H. Gamage Danula K, 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.

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.

Benson, A., D. Li, and K. Shue (2021). ""Potential" and the gender promotion gap". Mimeo.

Berk, J. B., C. R. Harvey, and D. Hirshleifer (2015). "Preparing a referee report: guidelines and perspectives". Mimeo.

Bhandari, M. *et al.* (2003). "Who did what? (Mis)perceptions about authors' contributions to scientific articles based on order of authorship". *Journal of Bone and Joint Surgery* 85 (8), pp. 1605–1609.

Bhandari, M. *et al.* (2004). "Interpreting authorship order and corresponding authorship". *Epidemiology* 15 (1), pp. 125–126.

Bhandari, M. *et al.* (2014). "Perceptions of authors' contributions are influenced by both byline order and designation of corresponding author". *Journal of Clinical Epidemiology* 67 (9), pp. 1049–1054.

Bohren, J. A., A. Imas, and M. Rosenberg (2019). "The dynamics of discrimination: theory and evidence". *American Economic Review* 109 (10), pp. 3395–3436.

Bransch, F. and M. Kvansnicka (2017). "Male gatekeepers: gender bias in the publishing process?" Mimeo.

Card, D. and S. DellaVigna (2013). “Nine facts about top journals in economics”. *Journal of Economic Literature* 51 (1), pp. 144–161.

Card, D. *et al.* (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.

Dehdari, S. H., E. Heikensten, and S. Isaksson (2019). “What goes around (sometimes) comes around: gender differences in retaliation”. Mimeo.

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.

Duch, J. *et al.* (2012). “The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact”. *PLOS ONE* 7 (12), e51332.

Ductor, L., S. Goyal, and A. Prummer (2018). “Gender and collaboration”. Mimeo.

Duffy, M. A. (2017). “Last and corresponding author practices in ecology”. *Ecology and Evolution* 7 (21), pp. 8876–8887.

Dworkin, J. D. *et al.* (2020). “The extent and drivers of gender imbalance in neuroscience reference lists”.

Ellison, G. (2002). “The slowdown of the economics publishing process”. *Journal of Political Economy* 10 (5), pp. 947–993.

Energy Economics (2021). *Full aims and scope.* (accessed: 2021-11-30).

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

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.

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.

Guarino, C. M. and V. M. H. Borden (2017). “Faculty service loads and gender: are women taking care of the academic family?” *Research in Higher Education* 58 (6), pp. 672–694.

Heckman, J. J. and S. Moktan (2019). “Publishing and promotion in economics: the tyranny of the Top Five”. *Journal of Economic Literature* (forthcoming).

Hengel, E. (2019). “Publishing while female. Are women held to higher standards? Evidence from peer review.” Mimeo.

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

Hillier, A. E. (2003). “Redlining and the home owners’ loan corporation”. *Journal of Urban History* 29 (4), pp. 394–420.

– (2005). “Residential security maps and neighborhood appraisals: the home owners’ loan corporation and the case of Philadelphia”. *Social Science History* 29 (2), pp. 207–233.

Hospido, L. and C. Sanz (2019). “Gender gaps in the evaluation of research: evidence from submissions to economics conferences”. Banco de España Working Paper No. 1918.

King, M. M. *et al.* (2017). “Men set their own cites high: gender and self-citation across fields and over time”. *Socius* 3, pp. 1–23.

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

Lang, K. and A. K.-L. Spitzer (2020). “Race Discrimination: An Economic Perspective”. *Journal of Economic Perspectives* 34 (2), pp. 68–89.

Lipman, B. (2009). “Why is language vague?” Mimeo.

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

Mattsson, P., C. J. Sundberg, and P. Laget (2011). "Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position". *Scientometrics* 87 (1), pp. 99–105.

Mayer, S. J. and J. M. K. Rathmann (2018). "How does research productivity relate to gender? Analyzing gender differences for multiple publication dimensions". *Scientometrics* 117 (3), pp. 1663–1693.

Means, M. L. and J. F. Voss (1996). "Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels". *Cognition and Instruction* 14 (2), pp. 139–178.

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

Phelps, E. S. (1972). "The statistical theory of racism and sexism". *American Economic Review* 62 (4), pp. 659–661.

Ratliff, K. A. and S. Oishi (2013). "Gender differences in implicit self-esteeem following a romantic partner's success or failure". *Journal of Personality and Social Psychology* 105 (4), pp. 688–702.

Rehg, M. T. *et al.* (2008). "Antecedents and outcomes of retaliation against whistleblowers: gender differences and power relationships". *Organization Science* 19 (2), pp. 221–240.

Reskin, B. (2012). "The race discrimination system". *Annual Review of Sociology* 38, pp. 17–24.

Sarsons, H. (2019). "Interpreting Signals in the Labor Market: Evidence from Medical Referrals". Mimeo.

Small, M. L. and D. Pager (2020). "Sociological perspectives on racial discrimination". *Journal of Economic Perspectives* 34 (2), pp. 49–67.

Symonds, M. R. E. *et al.* (2006). "Gender differences in publication output: towards an unbiased metric of research performance". *PLoS ONE* 1 (1), e127.

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.

West, J. D. *et al.* (2012). "The role of gender in scholarly authorship". *PLOS ONE* 8 (7), e66212.

Wren, J. D. *et al.* (2007). "The write position". *EMBO reports* 8 (11), pp. 988–991.

Appendices

A Descriptive statistics, by round

In Table A.1, we summarise basic statistics on the manuscripts in our sample. In the initial round of review (round 0), there are 2,356 manuscripts in our data, of which 14 percent have a female corresponding author.[1] Almost three-quarters of these manuscripts are asked to make major revisions at the end of the round; the remaining quarter are asked to make minor revisions. (Referees accept less than 1 percent of manuscripts in the very first round.) On average, papers spend about 5 months with referees and editors in the initial round of review. (Note that we do not observe "time spent revising" during the initial round of review.)

Table A.1: Descriptive statistics, by round

		Percentage of papers				Avg. no. days with	
Round	No. obs.	Female	Revise (major)	Revise (minor)	Accept	Refs.+Eds.	Authors
0	2,356	14.4	72.3	26.7	0.9	152.6	
1	2,336	14.6	12.0	36.5	51.5	57.9	97.5
2	1,136	16.9	4.4	19.6	75.9	22.5	32.6
3	274	15.7	4.0	16.5	79.5	12.9	22.5
4+	79	15.2	6.4	20.5	73.1	7.9	16.7

Note. Table displays round-specific descriptive statistics. The first column is is the round, the second column is the number of observations and the third column is the percentage of manuscripts with a female corresponding author. Columns 4–6 are the percentage of papers that received a "revise (major)", "revise (minor)" and "accept" decision at the end of the round. The final two columns show the average number of days a paper spent with referees/editors and with authors.

The second row of Table A.1 shows summary statistics on papers in the second round of review (round 1). The final column in Table A.1 suggests authors take, on average, about 3 months to revise their papers. (Recall that this round begins with the author revising her paper in response to referee reports received at the end of the previous round (see Figure 2).) Given so few manuscripts are immediately accepted after a single round of review, the number of observations in round 1 is very similar to the number of observations from round 0. The percentage of papers with a female corresponding author is also similar, but in contrast to the initial round, only a small proportion of papers are asked to make major revisions (12 percent). Most papers are accepted or asked to make only minor revisions, and the average number of days the manuscript spends with referees or editors is only about 2 months.

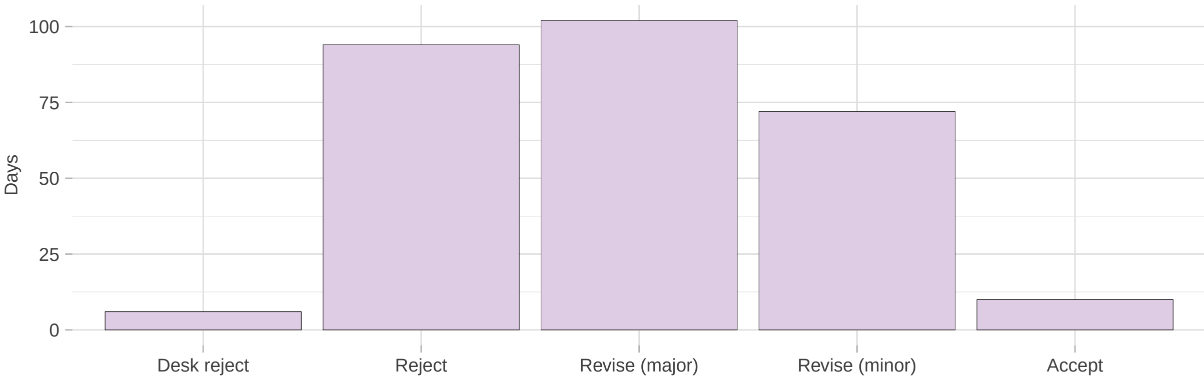
In round 2, the sample is roughly half the size as the sample from round 1, given just over 50 percent of papers were accepted in the previous round. Three quarters of papers are accepted this round; most of the rest are asked to make minor revisions. Authors also spend substantially less time revising, and referees spend less time reviewing compared to round 1. Consistent with Figure 1, the percentage of female-authored papers is two points higher than it was in the previous round. Relative to the round before it, conclusions are similar for rounds 3 and 4+.

[1] We do not nave data on every round of review for a very small number of submissions (see Footnote 2). As a result, our data include 2,359 submissions in total, but for three of those, data on the initial round of review is not available.

B Non-monotonic relationship between R_{ijt} and time$^{R}_{ijt}$

As we discuss in Section 3, one reason we restrict our analysis in this paper to accepted manuscripts is because referees appear to spend more time reviewing papers the *don't* reject than they spend reviewing papers they *do* reject, suggesting a non-monotonic relationship between R_{ijt} and evaluation time.

The evidence of this is presented in Figure B.1. It plots the median amount of time papers spend with referees by editor decision at the end of each round. (We report medians instead of averages in order to limit the impact of extreme outliers: a small number of papers appear to have been "forgotten" for several years before being eventually rejected—*e.g.*, one manuscript was under review for over seven years before being desk rejected.) Editors take a median of 6 days to reject papers they do not send out for review (desk reject); the median rejection delay for papers they do send out for review is 94 days. Among non-rejected papers, the medians for revise (major), revise (minor) and accept decisions are 102, 72 and 10 days, respectively.



Note. Graph plots the median amount of time spent with referees and editors by editor decision at the end of a round.

Figure B.1: Median amount of time spent with referees, by editor decision

The sample used to estimate Figure B.1 includes 13,351 unique manuscript-round instances, of which 3,755 were desk rejected manuscripts, 3,420 were manuscripts rejected after being sent out for review, 2,049 were "revise (major)" decisions, 1,766 were revise (minor) decisions and 2,361 were accept decisions. Note that these 2,361 accepted observations correspond to the same unique 2,359 full-length, regular issue manuscripts analysed in the main body of the paper. (Two manuscripts went through an additional round of review after being accepted.)

C Referee experience

Of the 4,721 unique referee-manuscript instances in our data, 82 percent are "novice" referees (*i.e.*, referees who had reviewed 1–4 manuscripts, including the current paper), 13 percent are "experienced" referees (5–19 papers), and the remaining 5 percent are "power" referees (20 or more papers). The top graph in Figure C.1 plots this histogram by author gender; the middle graph plots the distribution of the logarithm of 1 plus the number of papers previously reviewed. Referees with no prior experience reviewing for *Energy Economics* are slightly more likely to review male-authored papers; female-authored papers are slightly more likely to be reviewed by referees who had already reviewed 1–2 papers. Otherwise, both graphs suggest that men's and women's papers are reviewed by similarly experienced referees.

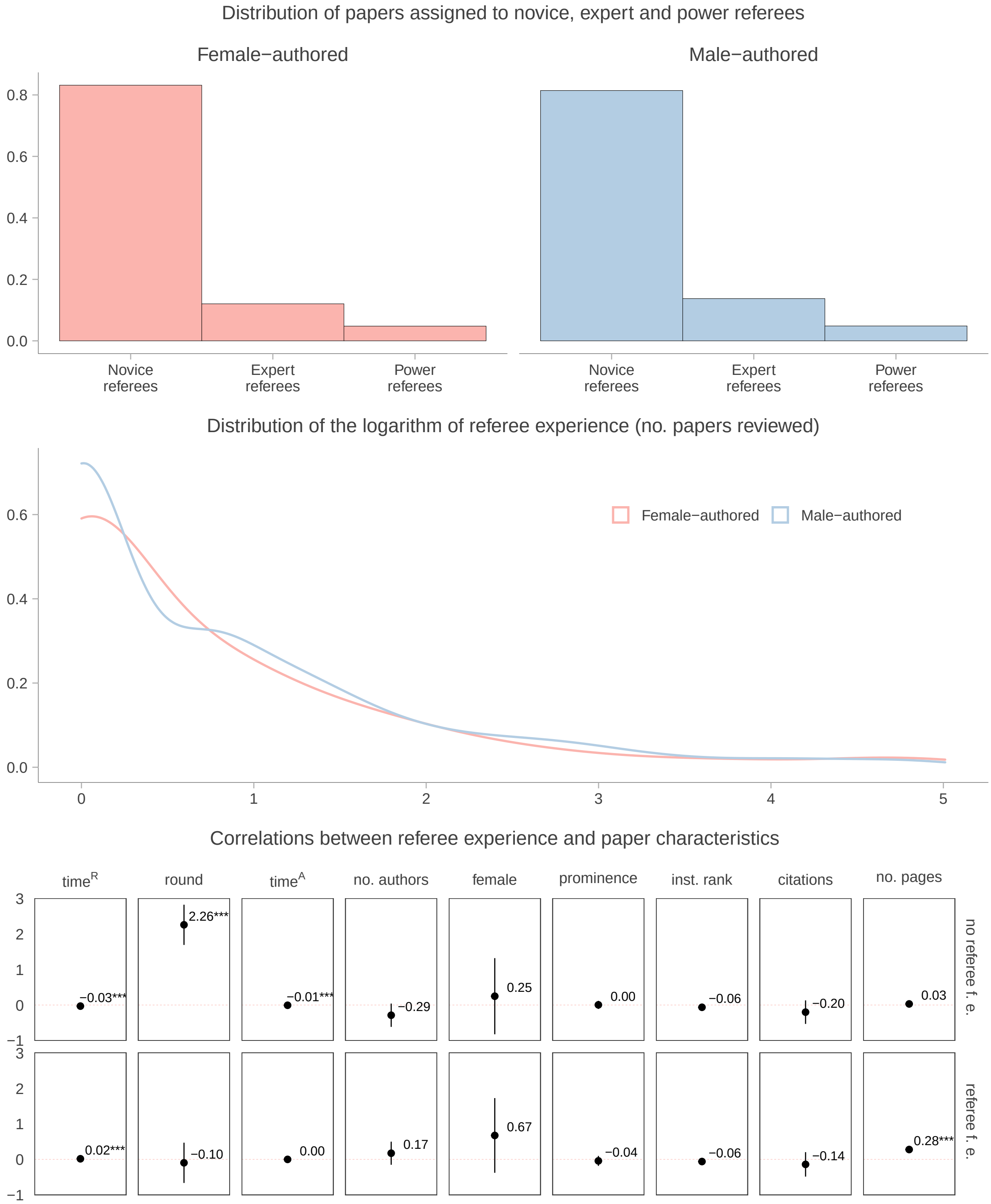
The bottom graph in Figure C.1 plots coefficients from nine separate OLS regressions of paper characteristics on referee experience, controlling for the year a manuscript was submitted. Without referee fixed effects (top row), more experienced referees are, on average, slightly quicker to return their reports. The manuscripts they review also go though more rounds and their authors spend less time revising. We do not, however, observe a statistically significant correlation between referee experience and a manuscript's page count, number of co-authors, citations (asinh) or its corresponding author's institutional rank, gender or prominence.[2]

With fixed effects (Figure C.1, bottom row), the coefficient on time^R reverses direction and is significant, suggesting that more experienced referees take slightly more time reviewing compared to their inexperienced selves. They papers they review are also longer, but the coefficients on round and time^A are no longer significant. The coefficients on remaining variables continue to be insignificant at traditional thresholds.

Identifying $\tilde{\Delta}^s$ in Equation (3) requires that referee assignment across author gender is orthogonal to the former's experience. This would be violated, for example, if editors assigned novice referees high quality male-authored papers and low quality female-authored papers and experienced referees the reverse. (In this example, any observed decline in gender reviewing and revising gaps would likely be due to the assignment mechanism itself rather than to referees' experience.) Although it is impossible to establish definitively that referee assignment is not related to author gender conditional on experience, we do not find obvious evidence of it. Table C.1 reports coefficients from an OLS regression of a manuscript's number of citations on the gender of its corresponding author, the experience of the referee and their interaction. Table C.2 repeats the exercise using corresponding author prominence as the dependent variable. Female-authored papers have lower citations and their authors are less prominent;[3] however, the interactions between experience and corresponding author gender are close to zero and statistically insignificant in all estimation models.

[2]Author prominence is the total number of papers the corresponding author previously published (as corresponding author) in *Energy Economics* at the time his manuscript is submitted. Results are similar if we proxy for author prominence using the maximum total number of papers any co-author had previously published in *Energy Economics*, regardless of corresponding author status (results available on request).

[3]Note that the coefficient on female in the first column declines to -0.038 (standard error 0.046) and is no longer statistically significant once the prominence of the corresponding author is controlled for.



Note. Top graph plots the distribution of papers assigned to novice (previously reviewed 1–4 papers, including the current paper), expert (5–19 papers) and "power" referees (20 or more papers). Middle graph plots the distribution of the logarithm of 1 plus the number of papers previously reviewed. Bottom graph plots coefficients from nine separate OLS regressions of paper characteristics on referee experience (number of papers previously reviewed), controlling for submission year fixed effects. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Figure C.1: Summary statistics on referee experience

Table C.1: Correlations between citations (asinh), referee experience and gender

	(1)	(2)	(3)	(4)	(5)
female (β_1)	−0.104**	0.001	−0.025	−0.034	−0.096
	(0.047)	(0.065)	(0.066)	(0.063)	(0.101)
experience	−0.001	−0.002	−0.002	−0.002	0.003**
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)
experience×female	0.001	0.002	0.002	0.002	0.000
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
No. obs.	4,721	4,721	4,721	4,721	4,721
R^2	0.441	0.450	0.456	0.510	0.792
Year	✓	✓	✓	✓	✓
Referee		✓	✓	✓	✓
Editor			✓	✓	✓
JEL (secondary)				✓	✓
Institution					✓

Note. Dependent variable is citation counts (asinh). Standard errors clustered on referees in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table C.2: Correlations between author prominence, referee experience and gender

	(1)	(2)	(3)	(4)	(5)
female (β_1)	−0.942***	−0.559***	−0.461***	−0.477***	−0.162
	(0.154)	(0.095)	(0.088)	(0.101)	(0.172)
experience	0.001	−0.002	−0.001	−0.005	−0.004
	(0.012)	(0.008)	(0.008)	(0.008)	(0.012)
experience×female	−0.005	0.000	−0.002	0.000	−0.002
	(0.013)	(0.005)	(0.005)	(0.005)	(0.006)
No. obs.	4,721	4,721	4,721	4,721	4,721
R^2	0.047	0.043	0.058	0.114	0.575
Year	✓	✓	✓	✓	✓
Referee		✓	✓	✓	✓
Editor			✓	✓	✓
JEL (secondary)				✓	✓
Institution					✓

Note. Dependent variable is corresponding author prominence (*i.e.*, the number of papers the corresponding author previously published in *Energy Economics* (as corresponding author) at the time a manuscript is submitted). Standard errors clustered on referees in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

D Cumulative gender gap

Tables D.1 and D.2 estimate the gender gap in the total number of days each referee spends reviewing a paper and each author spends revising his manuscript, respectively. Table D.3 estimates the gap using the difference (in days) between the dates papers were initially submitted and finally accepted as the dependent variable.

On average, papers go through about two rounds of review (see Section 2), so estimates in Tables D.1 and D.2 are unsurprisingly about twice the size of the estimates reported in Tables 1 and 2. According to Table D.3, female-authored papers spend a total of 4–6 weeks longer under review compared to male-authored papers with similar citations.

Table D.1: Gender gap in the total number of days spent with each referee

	(1)	(2)	(3)	(4)
female (β_1)	8.453***	8.602***	7.239**	12.527***
	(2.815)	(2.752)	(2.844)	(4.465)
citations (asinh)	−7.354***	−7.551***	−6.469***	−4.786***
	(1.323)	(1.339)	(1.165)	(1.445)
No. obs.	4,721	4,721	4,721	4,721
R^2	0.029	0.031	0.073	0.291
Year	✓	✓	✓	✓
Editor		✓	✓	✓
JEL (secondary)			✓	✓
Institution				✓

Note. Dependent variable is the total number of days (summed over all rounds) a referee spent reviewing a paper. Standard errors clustered by referee in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.2: Gender gap in the total number of days spent with authors

	(1)	(2)	(3)	(4)
female (β_1)	25.461***	23.205**	21.076**	32.243**
	(9.132)	(9.179)	(9.134)	(14.831)
citations (asinh)	−20.696***	−22.498***	−21.319***	−19.888***
	(3.154)	(3.103)	(3.163)	(5.325)
No. obs.	2,359	2,359	2,359	2,359
R^2	0.059	0.068	0.132	0.501
Year	✓	✓	✓	✓
Editor		✓	✓	✓
JEL (secondary)			✓	✓
Institution				✓

Note. Dependent variable is the total number of days (summed over all rounds) an author spent revising a paper. Standard errors clustered by author in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.3: Gender gap in the total number of days spent in review

	(1)	(2)	(3)	(4)
female (β_1)	37.746***	35.967***	29.634**	42.383**
	(12.058)	(12.136)	(12.378)	(20.206)
citations (asinh)	−56.764***	−55.521***	−53.020***	−46.036***
	(4.442)	(4.423)	(4.522)	(6.637)
No. obs.	2,366	2,366	2,366	2,366
R^2	0.137	0.143	0.214	0.569
Year	✓	✓	✓	✓
Editor		✓	✓	✓
JEL (secondary)			✓	✓
Institution				✓

Note. Dependent variable is the the difference (in days) between the dates papers were initially submitted to and finally accepted by *Energy Economics*. Robust standard errors in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

E Additional control variables

E.1 Author prominence

In this Appendix, we reproduce Tables 1 and 2 controlling for (i) the "prominence" of the corresponding author (Tables E.1 and E.2); and (ii) the prominence of the most prominent co-author (Tables E.3 and E.4). To proxy for "prominence", we total the number of papers an author previously published in *Energy Economics* at the time his manuscript is submitted. When controlling for corresponding author prominence, we only count papers on which an author was the corresponding author; when controlling for the prominence of the most prominent co-author, we count every paper he previously published in *Energy Economics* (whether he was corresponding author or not).

The gender gaps in reviewing and revising times are robust to controlling for author prominence. Without controls for referee fixed effects, results in Tables E.1 and E.3 suggest referees take 3.2–3.7 days longer to review female-authored papers. Conditional on referee fixed effects, however, the coefficient on female is identical to results reported in Table 1. Estimates and standard errors in Tables E.2 and E.4 are similar in magnitude to those reported in Table 2.

The coefficient on author prominence is consistently negative in columns (1)–(4) of each table, suggesting that referees are quicker to review papers by more prominent authors and more prominent authors are faster to revise their manuscripts. Once referee and/or institution fixed effects are account for, however, these conclusions no longer hold: referees appear to be quicker to review papers by less prominent authors, conditional on referee and author institution fixed effects and authors take about the same amount of time revising regardless of their prominence, conditional on author institution.

Table E.1: Gender gap in days spent with referees, controlling for corresponding author prominence

	Eq. 4				Eq. 5				Eq. 6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female (β_1)	3.382**	3.393**	3.436**	3.155**	3.394**	4.123**	3.270*	8.420***	9.889***
	(1.618)	(1.639)	(1.641)	(1.609)	(1.57)	(1.602)	(1.673)	(2.645)	(2.913)
t (round)	−16.012***	−13.274***	−14.325***	−13.232***	−15.096***	−13.329***	−13.222***	−13.531***	−13.572***
	(0.962)	(1.312)	(1.563)	(1.523)	(0.9)	(1.278)	(1.252)	(1.201)	(1.202)
corr. author prominence	−1.116***	−1.398***	−1.379***	−1.086***	−1.288***	−0.090	−0.070	0.464**	0.467**
	(0.212)	(0.209)	(0.212)	(0.207)	(0.215)	(0.217)	(0.215)	(0.227)	(0.226)
citations (asinh)	−4.564***			−4.561***	−3.824***	−2.950***	−3.110***	−2.912***	−3.024***
	(0.78)			(0.772)	(0.772)	(0.983)	(0.861)	(0.903)	(0.897)
R_{ijt} (referee's recommendation)									
revise (major)		7.734***		7.205***					
		(1.879)		(2.238)					
revise (minor)		6.580***		5.656***					
		(1.503)		(1.738)					
D_{it} (editor's decision)									
revise (major)			6.305**	1.503					
			(2.508)	(2.974)					
revise (minor)			5.229***	2.417					
			(1.673)	(1.918)					
Referee experience									
experience									0.217**
									(0.104)
experience×female									−0.105**
									(0.05)
No. obs.	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035
R^2	0.090	0.081	0.079	0.094	0.098	0.094	0.114	0.314	0.315
Oster bounds (β_1)	[2.1, 3.4]	[2.1, 3.4]	[2.2, 3.4]	[1.6, 3.2]	–	–	–	–	–
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
Referee						✓	✓	✓	✓
JEL (secondary)							✓	✓	✓
Institution								✓	✓

Note. Estimates are identical to those in Table 1 except that all results control for the prominence of the corresponding author at the time of submission (as measured by the number of previous publications in *Energy Economics*). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.2: Gender gap in days spent with authors, controlling for corresponding author prominence

	Eq. 7				Eq. 8				Eq. 9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	10.543**	10.541*	10.572*	9.698*	10.056*	10.166*	23.583***	19.773*	24.869**
	(5.268)	(5.449)	(5.458)	(5.292)	(5.316)	(5.27)	(8.033)	(10.77)	(10.836)
t	−41.911***	−29.029***	−38.740***	−28.408***	−42.576***	−43.554***	−45.079***	−58.883***	−58.834***
	(2.345)	(2.183)	(2.303)	(2.171)	(2.375)	(2.447)	(3.046)	(3.211)	(3.221)
corr. author prominence	−2.492***	−2.628***	−3.147***	−1.761***	−2.364***	−2.175***	0.048	0.991	0.968
	(0.398)	(0.535)	(0.426)	(0.484)	(0.424)	(0.572)	(0.986)	(0.769)	(0.772)
citations (asinh)	−11.365***			−11.619***	−12.024***	−11.509***	−12.621***	−12.561***	−12.457***
	(1.779)			(1.773)	(1.781)	(1.804)	(2.789)	(3.206)	(3.192)
D_{it-1} (revise (major))		45.450***		43.502***					
		(3.478)		(4.186)					
D_{it}									
revise (major)			24.769***	3.121					
			(7.082)	(7.401)					
revise (minor)			21.595***	5.742					
			(3.883)	(4.474)					
Referee experience									
experience									−0.042
									(0.166)
experience×female									−0.399**
									(0.19)
No. obs.	3,814	3,814	3,809	3,809	3,814	3,814	3,814	6,440	6,440
R^2	0.116	0.138	0.111	0.153	0.118	0.155	0.394	0.443	0.444
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
JEL (secondary)						✓	✓	✓	✓
Institution							✓	✓	✓
Referee								✓	✓

Note. Estimates are identical to those in Table 2 except that all results control for the prominence of the corresponding author (as measured by the number of previous publications in *Energy Economics*). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.3: Gender gap in days spent with referees, controlling for the prominence of the most prominent co-author

	Eq. 4				Eq. 5				Eq. 6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female (β_1)	3.698**	3.698**	3.705**	3.454**	3.739**	4.133***	3.277**	8.302***	9.784***
	(1.629)	(1.65)	(1.651)	(1.619)	(1.579)	(1.598)	(1.667)	(2.643)	(2.911)
t (round)	−15.967***	−13.238***	−14.121***	−13.066***	−15.075***	−13.327***	−13.222***	−13.519***	−13.559***
	(0.968)	(1.321)	(1.574)	(1.537)	(0.906)	(1.276)	(1.251)	(1.2)	(1.202)
max. co-author prominence	−0.695***	−0.953***	−0.949***	−0.666***	−0.795***	−0.106	−0.126	0.072	0.059
	(0.187)	(0.192)	(0.191)	(0.181)	(0.191)	(0.157)	(0.16)	(0.201)	(0.2)
citations (asinh)	−4.583***			−4.589***	−3.880***	−2.910***	−3.064***	−2.919***	−3.031***
	(0.78)			(0.774)	(0.775)	(0.969)	(0.854)	(0.902)	(0.895)
R_{ijt} (referee's recommendation)									
revise (major)		7.695***		6.818***					
		(1.884)		(2.233)					
revise (minor)		6.549***		5.419***					
		(1.5)		(1.746)					
D_{it} (editor's decision)									
revise (major)			6.811***	2.282					
			(2.522)	(2.985)					
revise (minor)			5.526***	2.809					
			(1.665)	(1.92)					
Referee experience									
experience									0.216**
									(0.108)
experience×female									−0.107**
									(0.051)
No. obs.	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035
R^2	0.087	0.079	0.077	0.091	0.095	0.094	0.114	0.313	0.315
Oster bounds (β_1)	[2.1, 3.4]	[2.1, 3.4]	[2.2, 3.4]	[1.6, 3.2]	–	–	–	–	–
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
Referee						✓	✓	✓	✓
JEL (secondary)							✓	✓	✓
Institution								✓	✓

Note. Estimates are identical to those in Table 1 except that all results control for the maximum number of papers previously published in *Energy Economics* at the time of submission across all co-authors on a paper. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.4: Gender gap in days spent with authors, controlling for the prominence of the most prominent co-author

	Eq. 7				Eq. 8				Eq. 9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	11.443**	11.404**	11.536**	10.409**	10.856**	10.987**	23.507***	19.640*	24.805**
	(5.247)	(5.417)	(5.418)	(5.269)	(5.294)	(5.243)	(8.02)	(10.727)	(10.825)
t	−41.775***	−28.805***	−38.561***	−28.156***	−42.500***	−43.524***	−45.099***	−58.875***	−58.825***
	(2.342)	(2.192)	(2.305)	(2.168)	(2.371)	(2.44)	(3.045)	(3.212)	(3.223)
max. co-author prominence	−1.271***	−1.477***	−1.868***	−0.765***	−1.182***	−0.970***	−0.268	0.377	0.394
	(0.365)	(0.278)	(0.351)	(0.267)	(0.348)	(0.317)	(0.665)	(0.562)	(0.563)
citations (asinh)	−11.573***			−11.856***	−12.302***	−11.880***	−12.609***	−12.558***	−12.453***
	(1.795)			(1.786)	(1.798)	(1.816)	(2.788)	(3.209)	(3.195)
D_{it-1} (revise (major))		45.828***		43.872***					
		(3.502)		(4.191)					
D_{it}									
revise (major)			24.874***	3.137					
			(7.105)	(7.404)					
revise (minor)			22.053***	5.931					
			(3.902)	(4.478)					
Referee experience									
experience									−0.044
									(0.164)
experience×female									−0.404**
									(0.19)
No. obs.	3,814	3,814	3,809	3,809	3,814	3,814	3,814	6,440	6,440
R^2	0.114	0.136	0.109	0.152	0.116	0.154	0.394	0.443	0.444
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
JEL (secondary)						✓	✓	✓	✓
Institution							✓	✓	✓
Referee								✓	✓

Note. Estimates are identical to those in Table 2 except that all results control for the maximum number of papers previously published in *Energy Economics* at the time of submission across all co-authors on a paper. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

E.2 Manuscript length

In Tables E.5 and E.6 we reproduce Tables 1 and 2 controlling for the number of pages in the final published version of a paper. The coefficients on female in both tables are very similar to those reported in Tables 1 and 2.

In Table E.6, the coefficient on manuscript length is consistently positive and (usually) at least weakly significant. It suggests that for each additional page in a manuscript, authors take up to two days longer revising their papers.

The coefficient on manuscript length is also positive in Table E.5, columns (1)–(5); however, it loses significance once referee fixed effects are accounted for. In other words, referees spend more time reviewing longer papers, but this appears to be due to an underlying correlation between the assignment of referees and manuscript length, since the same referee takes about the same amount of time to review a paper regardless of its length.

Table E.5: Gender gap in days spent with referees, controlling for manuscript length

	Eq. 4				Eq. 5				Eq. 6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female (β_1)	4.163**	4.582***	4.558***	3.869**	4.274***	4.169***	3.332**	8.428***	9.951***
	(1.68)	(1.723)	(1.718)	(1.66)	(1.615)	(1.595)	(1.664)	(2.652)	(2.925)
t (round)	−15.933***	−13.010***	−13.650***	−12.838***	−15.093***	−13.332***	−13.224***	−13.507***	−13.547***
	(0.971)	(1.346)	(1.642)	(1.574)	(0.903)	(1.278)	(1.253)	(1.197)	(1.199)
no. pages	0.518***	0.382*	0.381*	0.486**	0.548***	0.001	−0.048	−0.255	−0.308
	(0.195)	(0.208)	(0.207)	(0.196)	(0.208)	(0.176)	(0.186)	(0.277)	(0.282)
citations (asinh)	−5.369***			−5.334***	−4.835***	−2.973***	−3.119***	−2.871***	−2.977***
	(0.825)			(0.821)	(0.834)	(0.991)	(0.866)	(0.891)	(0.885)
R_{ijt} (referee's recommendation)									
revise (major)		8.350***		6.694***					
		(1.962)		(2.235)					
revise (minor)		6.451***		5.163***					
		(1.509)		(1.761)					
D_{it} (editor's decision)									
revise (major)			7.848***	2.970					
			(2.696)	(3.114)					
revise (minor)			5.534***	2.894					
			(1.702)	(1.954)					
Referee experience									
experience									0.223**
									(0.109)
experience×female									−0.107**
									(0.051)
No. obs.	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035
R^2	0.085	0.071	0.070	0.089	0.091	0.093	0.114	0.313	0.315
Oster bounds (β_1)	[3.7, 4.2]	[4.5, 4.6]	[4.5, 4.6]	[3.1, 3.9]	–	–	–	–	–
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
Referee						✓	✓	✓	✓
JEL (secondary)							✓	✓	✓
Institution								✓	✓

Note. Estimates are identical to those in Table 1 except that all results control for the length of the published manuscript. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.6: Gender gap in days spent with authors, controlling for manuscript length

	Eq. 7				Eq. 8				Eq. 9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	11.949**	12.225**	12.722**	10.581**	11.325**	11.020**	22.097***	18.254*	23.382**
	(5.302)	(5.509)	(5.571)	(5.274)	(5.329)	(5.242)	(7.91)	(10.68)	(10.799)
t	−41.734***	−28.364***	−38.325***	−28.040***	−42.529***	−43.654***	−45.630***	−58.971***	−58.918***
	(2.412)	(2.323)	(2.416)	(2.248)	(2.418)	(2.463)	(3.06)	(3.22)	(3.233)
no. pages	1.007**	0.796*	0.629	1.011**	1.014**	1.166**	2.303***	1.371*	1.416**
	(0.444)	(0.426)	(0.449)	(0.408)	(0.459)	(0.463)	(0.68)	(0.712)	(0.717)
citations (asinh)	−12.679***			−12.611***	−13.421***	−12.792***	−13.027***	−12.773***	−12.660***
	(1.857)			(1.775)	(1.878)	(1.871)	(2.795)	(3.206)	(3.192)
D_{it-1} (revise (major))		47.038***		44.762***					
		(3.701)		(4.216)					
D_{it}									
revise (major)			25.485***	2.587					
			(7.225)	(7.393)					
revise (minor)			22.275***	5.411					
			(4.007)	(4.481)					
Referee experience									
experience									−0.074
									(0.168)
experience×female									−0.403**
									(0.189)
No. obs.	3,814	3,814	3,809	3,809	3,814	3,814	3,814	6,440	6,440
R^2	0.113	0.134	0.105	0.152	0.115	0.154	0.397	0.444	0.444
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
JEL (secondary)						✓	✓	✓	✓
Institution							✓	✓	✓
Referee								✓	✓

Note. Estimates are identical to those in Table 2 except that all results control for the length of the published manuscript. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

E.3 Number of co-authors

In Tables E.7 and E.8 we reproduce Tables 1 and 2 controlling for the number of co-authors on a paper. The coefficients on female in both tables are very similar to those reported in Tables 1 and 2. The number of co-authors on a paper appears to be weakly positive or not related to the gender gap in days spent with referees (Table E.7). Authors appear to spend 2–3 extra days revising for every co-author on a paper, but these estimates are either only weakly significant or not significant (Table E.8).

Table E.7: Gender gap in days spent with referees, controlling for number of co-authors

	Eq. 4				Eq. 5				Eq. 6
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female (β_1)	4.466***	4.789***	4.760***	4.146**	4.475***	4.169***	3.301**	8.353***	9.854***
	(1.659)	(1.698)	(1.693)	(1.639)	(1.602)	(1.594)	(1.667)	(2.637)	(2.904)
t (round)	−15.889***	−12.915***	−13.554***	−12.698***	−15.036***	−13.332***	−13.223***	−13.519***	−13.560***
	(0.973)	(1.343)	(1.632)	(1.566)	(0.907)	(1.278)	(1.253)	(1.201)	(1.203)
no. co-authors	0.967*	0.618	0.563	0.984*	1.026**	0.004	0.151	−0.354	−0.404
	(0.506)	(0.506)	(0.509)	(0.507)	(0.503)	(0.503)	(0.535)	(0.668)	(0.664)
citations (asinh)	−5.317***			−5.291***	−4.772***	−2.973***	−3.141***	−2.895***	−3.003***
	(0.843)			(0.839)	(0.854)	(1.007)	(0.887)	(0.902)	(0.896)
R_{ijt} (referee's recommendation)									
revise (major)		8.537***		6.889***					
		(1.951)		(2.234)					
revise (minor)		6.574***		5.258***					
		(1.512)		(1.757)					
D_{it} (editor's decision)									
revise (major)			8.076***	3.116					
			(2.662)	(3.075)					
revise (minor)			5.712***	3.049					
			(1.7)	(1.944)					
Referee experience									
experience									0.217**
									(0.109)
experience×female									−0.107**
									(0.051)
No. obs.	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035	7,035
R^2	0.084	0.071	0.069	0.088	0.090	0.093	0.114	0.313	0.315
Oster bounds (β_1)	[3.7, 4.2]	[4.5, 4.6]	[4.5, 4.6]	[3.1, 3.9]	–	–	–	–	–
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
Referee						✓	✓	✓	✓
JEL (secondary)							✓	✓	✓
Institution								✓	✓

Note. Estimates are identical to those in Table 1 except that all results control for number of co-authors. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table E.8: Gender gap in days spent with authors, controlling for number of co-authors

	Eq. 7				Eq. 8				Eq. 9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
female	12.383**	12.558**	12.975**	11.001**	11.555**	11.507**	23.402***	19.057*	24.155**
	(5.331)	(5.523)	(5.577)	(5.311)	(5.375)	(5.3)	(8.032)	(10.816)	(10.915)
t	−41.462***	−28.143***	−38.119***	−27.732***	−42.257***	−43.329***	−45.027***	−58.828***	−58.777***
	(2.382)	(2.291)	(2.377)	(2.216)	(2.395)	(2.445)	(3.049)	(3.21)	(3.221)
no. co-authors	2.781*	1.963	1.665	3.062*	2.886*	2.651*	2.055	2.985	2.992
	(1.577)	(1.599)	(1.602)	(1.57)	(1.586)	(1.586)	(2.115)	(1.955)	(1.954)
citations (asinh)	−12.705***			−12.665***	−13.448***	−12.744***	−12.773***	−12.787***	−12.680***
	(1.857)			(1.771)	(1.866)	(1.856)	(2.804)	(3.198)	(3.186)
D_{it-1} (revise (major))		47.080***		44.615***					
		(3.684)		(4.217)					
D_{it}									
revise (major)			25.864***	3.290					
			(7.225)	(7.407)					
revise (minor)			22.485***	5.800					
			(3.984)	(4.486)					
Referee experience									
experience									−0.051
									(0.163)
experience×female									−0.399**
									(0.191)
No. obs.	3,814	3,814	3,809	3,809	3,814	3,814	3,814	6,440	6,440
R^2	0.113	0.134	0.105	0.152	0.115	0.153	0.394	0.443	0.444
Year	✓	✓	✓	✓	✓	✓	✓	✓	✓
Editor					✓	✓	✓	✓	✓
JEL (secondary)						✓	✓	✓	✓
Institution							✓	✓	✓
Referee								✓	✓

Note. Estimates are identical to those in Table 2 except that all results control for the number of co-authors on a paper. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.