Gender and the time cost of peer review*

Diane Alexander† Olga Gorelkina‡ Erin Hengel§ Richard Tol¶

April 2023
(First version: 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 publishing: length of peer review. Using detailed administrative data from an economics field journal, we find that, conditional on manuscript quality, referees spend longer reviewing female-authored papers, are slower to recommend accepting them, manuscripts by women go through more rounds of review and their authors spend longer revising them. Less disaggregated data from 32 economics and finance journals corroborate these results. We conclude by showing that all gender gaps decline—and eventually disappear—as the same referee reviews more papers. This pattern suggests novice referees initially statistically discriminate against female authors, but are less likely to do so as their information about and confidence in the peer review process improves. More generally, they also suggest that women may be particularly disadvantaged when evaluators are less familiar with the objectives and parameters of an assessment framework.

Keywords: Gender Inequality, Statistical Discrimination, Research Productivity, Peer Review; *JEL*: A11, D8, J16, J24, J7.

*We are especially grateful to Amanda McFarland and Shangen Lu for excellent research assistance. We also appreciate helpful comments from Liam Beiser-McGrath, Thomas Biegert, Pilar Cuevas-Ruiz, Francesca Foliano, Marlène Koffi, Alexander Mcquoid, Chiara Orsini, Eva Ranehill, Hannes Schwandt, Almudena Sevilla, Sarah Smith, Margaux Suteau, seminar participants at the University of Liverpool, American University, Queen Mary University, Toulouse School of Economics, Stockholm University (SOFI), University of Kent, University of Sussex, University of Bath, University of Sheffield, Brunel University, Discrimination and Disparities Seminar, Webinar in Gender, and Family Economics, FemQuant Seminar and conference participants at the Royal Economic Society Annual Conference, Conference on Diversity, Equity and Inclusion in Economics, Finance and Central Banking, Philosophy of Science Biennial Meeting, EEA-ESEM Annual Conference, URPP Conference on Gender and Inequality, Pathways to Gender Equality in Central Banks Conference, Scottish Economic Society Annual Conference, and Eastern Economic Association Annual Meeting.

†The Wharton School, dealexa@wharton.upenn.edu.

‡University of Liverpool, ogorelkina@gmail.com.

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

¶University of Sussex, Vrije Universiteit Amsterdam, Tinbergen Institute, CESifo, Colorado School of Mines and Abu Dhabi University, rtol@sussex.ac.uk.

1 Introduction

On average, men in academia publish more papers than women (West *et al.* 2012). The gap remains large after controlling for observable factors (Ductor *et al.* 2021). It also emerges surprisingly early—women take five years to reach the same annual rate of output that men achieve in the first two years of their careers (Symonds *et al.* 2006).

Yet this gender "productivity gap" is not identical across fields, 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 for papers published in less prestigious outlets and in fields where research is cheaper to produce (Duch *et al.* 2012; Mayer and Rathmann 2018). It may also reverse 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 explore one potential explanation for the gender productivity gap: whether women experience longer peer review. If female-authored manuscripts take longer to publish, then women will have less time and resources for starting new projects. As a result, they may struggle to meet strict publication requirements within fixed deadlines (*e.g.*, for tenure or promotion), submit competitive grant funding applications (particularly earlier in their careers), and secure permanent employment in research-intensive institutions. All of these factors likely contribute to slower publication rates compared to men.

To investigate this issue, we ask the following five questions: (i) do referees review papers by female authors as quickly as they review papers of similar quality by male authors? (ii) are referees as fast to accept female-authored papers? (iii) do female-authored papers go through more rounds of review? (iv) do women exert more effort responding to referees? and (v) to what extent do these gaps depend on how well-informed referees are about the peer review process at particular journals?

To answer these questions, we construct two unique administrative datasets. Our first dataset contains comprehensive information on every round of the review process for articles published in the top field journal *Energy Economics*—including dates authors submitted (or re-submitted) their papers, dates referees agreed to review and eventually returned their reports, referees' and editors' round-specific decisions, and unique identifiers that track referees and authors as they review or submit multiple manuscripts. Our second dataset contains similar—but more limited—information on the length of peer review for 32 economics and finance journals published by Elsevier. We augment both datasets with additional bibliographic and demographic data, including citation counts, secondary *JEL* codes, and authors' genders and institutions.

Conditional on manuscript quality, referees at *Energy Economics* take longer reviewing and accepting female-authored papers, female-authored papers go through more rounds of review, and their authors spend longer revising them. More specifically, in every round of review, referees spend 4–10 more days reviewing papers by female corresponding or solo authors; they are also 1–3 percent less likely to recommend accepting them. Compared to similar quality male-authored papers, female-authored papers go through 0.07–0.1 more rounds of review and, in each round, their authors spend 11–31 more days revising them. These gaps are robust to controlling for editors, referees, secondary *JEL* codes, authors' institutional rank, author prominence, manuscript length, and number of co-authors. They also replicate using alternative definitions of female authorship (including exclusively female, senior female and the ratio of female authors).[1]

We find similar gender gaps in time and length of review when we analyse the more aggregated data for 32 economics and finance journals. Conditional on quality, corresponding and solo female-authored

[1] Coefficient magnitudes are smaller and standard errors higher when female authorship is defined as having 50 percent or more female authors or at least one female author (see Appendix A.5).

manuscripts go through 0.04–0.08 additional rounds of review; their authors also spend 6–22 days longer revising them. Unfortunately, these data do not contain round-by-round information on time spent with each referee; to approximate it, we use the difference between a paper's submission and first decision dates. In contrast to results for *Energy Economics*, we find that referees spend about the same amount of time evaluating female-authored papers; however, an initial gender gap in referee time appears to be masked by female-authored papers that are subsequently subject to particularly tough peer review. When we restrict the sample to papers that were accepted after a single round (the most likely outcome), the gender gap is almost identical to the one we found for *Energy Economics* (4–8 days).

As we illustrate in Appendix D, these gender gaps are economically significant. The cumulative time gap is 29–60 days at *Energy Economics* and 18–29 days at the other 32 economics and finance journals. A back-of-the-envelope calculation suggests that for every 10 papers a woman publishes, she will have spent 6–20 months longer under review compared to a man with similar research quality and quantity.

As discussed in Section 3.1, we normatively interpret these gaps as non-desirable and the result of direct or systemic taste-based discrimination, disparate treatment resulting from power differentials or statistical discrimination, where we broadly define the latter to include any kind of differential treatment that would not have occurred if information were perfect. Nevertheless, this interpretation requires making several strong assumptions. First, our proxies for the selection criteria at *Energy Economics* must not be biased in favour of women. Although it is impossible to entirely rule out the converse, a large body of research consistently suggests that our main proxy for quality—citations—probably isn't (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Ferber 1988; Koffi 2019). We also find similar results when controlling for alternative proxies, including referees' and editors' round-specific decisions and editors' decisions in the previous and first rounds.

Second, normatively interpreting non-zero gender gaps as evidence of direct or systemic discrimination implicitly assumes that these gaps are not counterbalanced by factors outside the narrow system we study. For example, suppose male-authored papers are rejected more often. Then men may spend less time under review at particular journals, but more time across all journals such that their total time in review equals women's over the lifespan of a manuscript. Although answering this question is beyond the scope of our paper, we note that existing evidence does not suggest that male–authored papers are rejected at higher rates, conditional on journal and selection criteria (see, for example, Card *et al.* 2020).

Finally, we do not observe authors' effort revising their papers. Instead, we imperfectly proxy for it using the time they spend responding to referees, which undoubtedly contains substantial non-classical measurement error—for example, greater childcare obligations, teaching loads or service responsibilities may force women to push their eventual resubmission dates further into the future, holding effort fixed. Because we have no reliable means of accounting for these factors, we encourage additional caution when interpreting these particular results.

Having established clear evidence of gender gaps in the length of peer review, we next study how direct or systemic statistical discrimination, broadly defined, contributes to their formation. To do so, we exploit the insight that better informed referees will be more knowledgeable about whether a paper meets the standards of acceptance at *Energy Economics*. In addition, the signal they receive from a paper about its quality will also be more accurate—and contain fewer gender-specific distortions—than the signal received by less informed referees. As a result, better informed referees should be quicker to review and accept manuscripts, all else equal; clearer signals and greater knowledge should also boost their ability to write less ambiguous, easier-to-implement reports.

We find that referees' knowledge and information about the peer review process at *Energy Economics*—which we proxy for using their experience refereeing—plays a pivotal role in diminishing gender gaps in the length of peer review. The magnitude of every gender gap is greater when papers are evaluated by poorly informed referees; when reviewed by well-informed referees, however, all four gaps decline and possibly reverse. This pattern suggests that novice referees initially statistically discriminate against female authors, but as they gain experience, their ability to identify quality in men's and women's papers converges, so gaps in reviewing time, acceptance rates and rounds fall. Although revision time remains an imperfect proxy for effort, its similar decline nevertheless indicates that experienced referees write reports that female authors are especially quick to respond to. Moreover, if experienced referees are not biased *against* men (conditional on selection criteria), then the extra time women spend responding to novice referees relative to expert referees compared to men—which we estimate is between 69–106 days—conservatively approximates the cost to women of statistical discrimination.

These results hold after accounting for referee fixed effects, suggesting that gender gaps decline even as the *same* referee reviews more papers. They are also robust to controlling for editors, secondary *JEL* codes, institutional rank, author prominence, number of co-authors and manuscript length. We caution, however, that better information causes the gaps to decline only if poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the same manuscripts, an assumption we explore more carefully in Appendix A.4.

Identifying the pivotal role played by referee experience also helps us pinpoint potential solutions to the gendered cost of peer review. As referees gain knowledge about the process of peer review at *Energy Economics*, gender gaps in review length decline, suggesting that editors may be able to increase equity in outcomes—while holding the informativeness of the refereeing process constant—by expanding the pool of competent and experienced referees and assigning them more female-authored manuscripts.

This paper contributes to numerous strands of the literature. First, we join several studies suggesting that discriminatory selection procedures may be largely driven by evaluators with low quality information or a lack of experience. In particular, Brock and De Haas (2022) use a static setting to show that younger, less experienced loan officers are more likely to require a guarantor from female applicants compared to otherwise identical male applicants. Within a dynamic framework, Bohren *et al.* (2019) show that evaluators are less likely to discriminate against women as they learn more about their skills. Like Bohren *et al.* (2019), we observe how outcomes change as the same evaluator is exposed to better information, but similar to the former, we focus on evaluators' knowledge about the process and standards of assessment, holding the information they have about the individuals they are evaluating fixed. This allows us to identify a passive—and likely more feasible—means of using information to quickly mitigate the negative consequences of bias: rather than giving reviewers more information about each person they are evaluating, assign more experienced reviewers to evaluate women.

Second, we provide further evidence that women receive lower quality, harder-to-implement and/or less relevant advice compared to men. For example, Gallen and Wasserman (2022) find that female students receive substantially more information on work/life balance when seeking career advice from a mentor, and this may deter them from their preferred career path. Evidence in Bucher-Koenen *et al.* (2021) suggests that financial advisors offer more self-serving and financially costly advice to women, thus lowering their investment returns relative to men's. In our study, we find that women spend relatively more time revising when they are "treated" with less experienced referees. We interpret this result as evidence that inexperienced referees view women's papers as "riskier"—possibly due to stereotyping or because they are less familiar with female-authored work—and "hedge" their positive decisions by writing tougher reports.

Third, our study complements research exploring how uncertainty creates and exacerbates gender outcome gaps. For example, Bowles *et al.* (2005) find that women perform better in negotiations with less situational ambiguity. Similarly, greater clarity on required qualifications and less uncertainty surrounding expected on-the-job effectiveness increase women's and men's application rates to male- and female-typed job opportunities, respectively (Coffman *et al.* 2023; Delfino 2022). We add to this research by studying ambiguity on the other side of the assessment process and finding that gendered outcomes are also more common when evaluators are less familiar with the objectives and standards of selection.

Finally, our evidence suggests that peer review may impose greater costs on female researchers (here economists) wherever they submit, thus providing further evidence that women are often subject to tougher standards than men (see *e.g.*, Card *et al.* 2020; Hengel 2022; Hengel and Moon 2022). Most relevant to our work, Hengel (2022) shows that female-authored papers published in two top general interest economics journals spend three to six months longer under review compared to observably equivalent male-authored papers. In addition to finding similar top-line patterns in a wide selection of journals, we build on this work by using more detailed data to investigate the underlying mechanisms driving these gaps.

This paper proceeds in the following order. Section 2 describes the data. Section 3 estimates gender gaps in time and length of peer review. Section 4 identifies the role of information in creating and mitigating these gaps. Section 5 concludes.

2 Data

2.1 *Energy Economics*

Our primary 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 top field journal focused on the areas of energy economics and energy finance; in 2020, its Clarivate Analytics Journal Impact Factor and Scopus CiteScore were 7.042 (2 out of 376 economics journals) and 10.0 (22 out of 661 economics journals), respectively.[2]

Papers submitted to *Energy Economics* are either desk-rejected by the editor or undergo single-blind peer review. Although submitting authors are asked to recommend referees, the handling editor ultimately decides whom to approach for reviews. Unlike many other economics journals, *Energy Economics* has a stated policy of rejecting papers unless they are favourably reviewed by at least two referees.

Unless otherwise mentioned, we study eventually accepted manuscripts, only. One of our principal objectives is to better understand how differences in the (perceived) advice men and women receive can potentially lead to (real) gender differences in subsequent productivity. As is well established in the "cheap talk" literature, however, allowing players to exchange costless, non-verifiable messages can generate so-called "babbling" equilibria—*i.e.*, situations in which experts convey uninformative messages that decision makers consequently ignore. To avoid this outcome, we restrict our analysis to non-rejected papers. Conditional on recommending acceptance or revise and resubmit, referees' interests will be better aligned with authors'; as a result, their communication should be more meaningful and "co-ordinated" (Matthews *et al.* 1991).[3]

We also have a practical reason for focusing on eventually accepted papers: a handful of international programmes require students to submit their master's theses as a condition of the degree. This policy leads to a large number of rejected submissions for which we cannot easily assign gender nor obtain any kind of externally decided proxy of quality (*e.g.*, citations).[4]

[2]A recent study suggests that the authors and editors at *Energy Economics* are particularly geographically diverse (Angus *et al.* 2021).

[3]Consistent with this hypothesis, we find that referees spend more time reviewing papers they don't reject than they spend reviewing papers they do reject (see Appendix A.1).

[4]Many of these papers are never ultimately published, and their authors have left academia and so do not maintain a personal website or post their CVs and papers online.

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 the second step, we collected publicly available information on manuscripts (*e.g.*, Web of Science citation counts) and identified referees and authors with multiple EEM accounts. To proxy for a manuscript's "gender", we primarily 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). We also restrict the sample to solo-authored papers and, in Appendix A.5, report results using alternative proxies, including exclusively female-authored, senior female-authored and the ratio of female authors. Genders for all authors 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 individuals; and (iv) by contacting authors themselves or people and institutions connected to them. We identified the genders of 99.7 percent of all authors in the data.

In the final step, we extracted 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. We then merged these data with the data collected in steps 1 and 2 and anonymised referees', editors' and authors' identities. Our final dataset includes 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. Of these papers, 2,016 have a male corresponding author, 343 have a female corresponding author and 447 were solo-authored (387 by men and 60 by women).[5]

2.1.1 Summary statistics

The percentage of female-authored papers published in *Energy Economics* has not changed much over time (Figure 1A): in 2006, 14.3 percent of published articles had a female corresponding author, compared to 14.5 percent in 2018. As Figure 1B illustrates, however, the total 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*).

Female-authored papers are more likely to go through more rounds of review before publication (Figure 1C), though the average number of rounds is steady over our sample period (Figure 1D).[6] According to Figure 1C, 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.[7]

Female-authored papers also spend more time with referees. Figure 1E displays the distribution of time spent with referees by manuscript gender. On average, male- and female-authored papers spend 220 and 238 days with referees, respectively. Figure 1F shows that referees spent slightly more time reviewing

[5] According to RePEc (2023), 24 percent of energy economists and 26 percent of all economists are female.

[6] Selection bias drives 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. Manuscripts included in our data must have been submitted on or after April 2005 and published on or before June 2019; thus, papers published in 2006 or submitted in 2018 would have necessarily experienced a faster than average review process.

[7] A Kolmogorov-Smirnov test suggests that the two samples are not drawn from the same probability distribution ($D = 0.09$, p-value $= 0$).

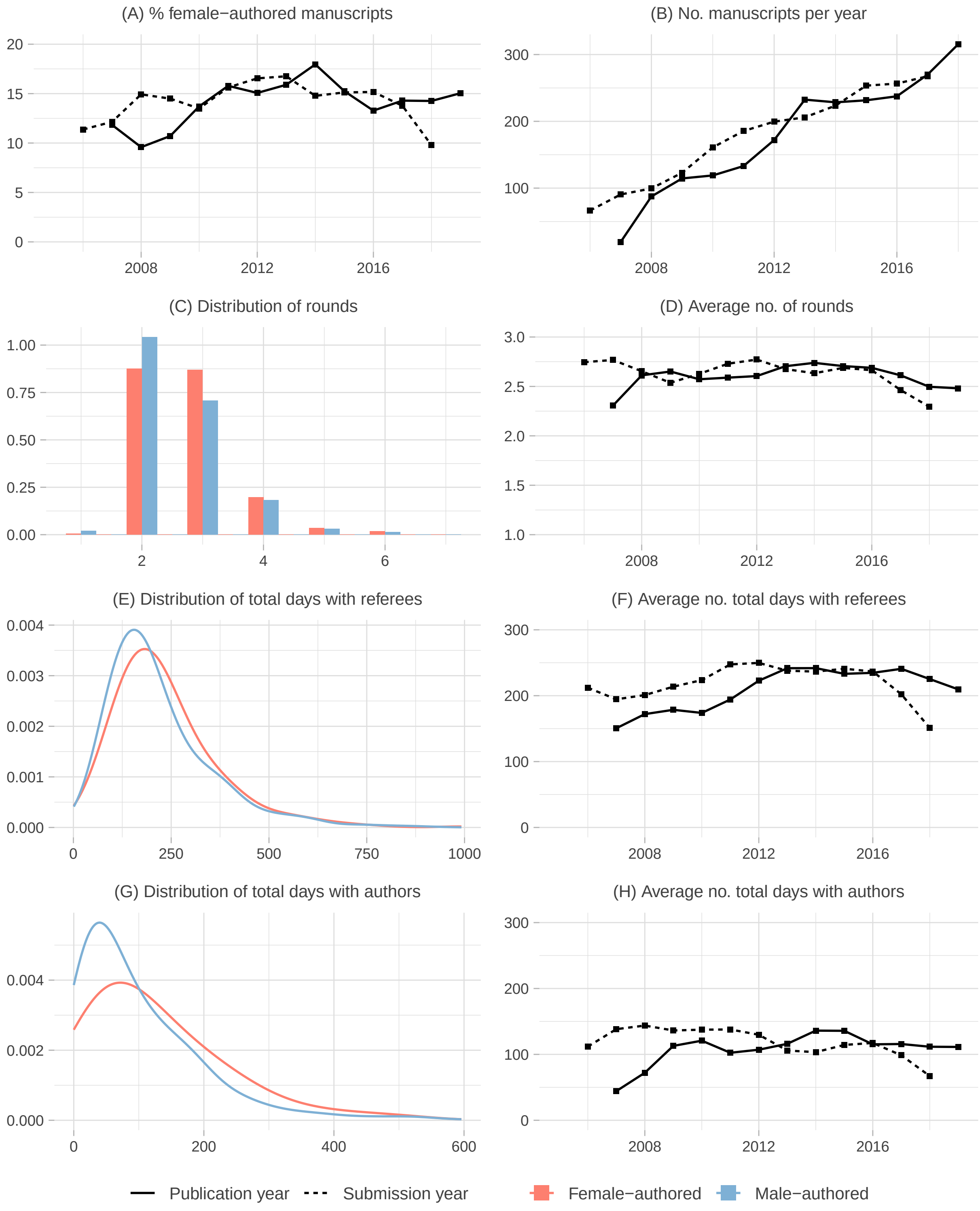

Figure 1: Summary statistics for manuscripts published in *Energy Economics*

Note. Graph (A) is the two-year moving average of the percentage of papers published and submitted each year that are female-authored, where "female-authored" refers to papers with a female corresponding author; Graph (B) displays the two-year moving average of the total number of manuscripts published and submitted each year. Graphs (C) and (D) plot the distribution and two-year moving average of total number of rounds, respectively. Graphs (E) and (F) plot the distribution and two-year moving average of the total number of days a manuscript spends with referees, respectively. Graphs (G) and (H) show the distribution and the two-year moving average of the total number of days manuscripts spend with authors, respectively.

papers in 2010 than they did in 2005, but between 2010 and 2019 the average number of days has been relatively flat.

Finally, female authors spend more time revising than do male authors. Figure 1G 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. Yet the combination of growing revision times and female representation does not appear to drive this pattern. 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 (Figure 1H).

In Appendix A.2, we summarise additional 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 with round, but female-authorship positively correlates with it. This later observation is consistent with the evidence in Figure 1 suggesting that female-authored papers go through more rounds of review.

2.2 32 economics and finance journals

Our secondary dataset contains more limited information on the time and length of peer review for 32 economics and finance journals published by Elsevier.[8] Elsevier obtained permission from all journals' editorial boards and extracted the data in 2019 themselves. As a condition of using the data, we agreed not to release results that could identify gender gaps at specific journals.

The full dataset contains the following information on 164,809 manuscripts: journal name, basic data on the submitting author (title, first name, country of residence and predicted gender), final editorial decision, total rounds of review, and submission, first decision, final acceptance and publication dates. We excluded 138,003 manuscripts that were rejected, withdrawn by the authors, removed by another party (*e.g.*, a journal editor or Elsevier administrator), immediately "desk accepted" upon submission or published in a special issue.[9]

To obtain additional information on manuscripts—including citation counts and authors' full names—we matched each one to an article published in the same journal using submitting author's first names and submission, acceptance and publication dates. (This information is almost always published in the typeset version of accepted papers.) Matches were manually verified in all instances where a manuscript matched to more than one published article, a published article matched to more than one manuscript, two or more dates did not match or no co-author's first name in the published article exactly matched the first name of the author who submitted the manuscript. In total, we matched 98 percent of accepted manuscripts.

In the data provided by Elsevier, each submitting author had been assigned a gender (male, female or unknown) using a combination of non-gender-neutral titles (*e.g.*, "Ms." or "Mr.") and country-specific name lists. We subsequently manually verified the genders of female submitting authors and attempted to assign genders to submitting authors with an unknown gender. Our final dataset includes 24,560 accepted manuscripts, 3,980 with a female submitting author, 20,580 with a male submitting author and 6,553 solo-authored papers (1,008 by women and 5,545 by men). Unless otherwise mentioned, we exclude the 480 papers submitted by an author of unknown gender. Appendix B.2 presents summary statistics, most of which resemble the graphs shown in Figure 1.

[8] Data from a 33rd journal, *International Economics*, are excluded because it is not indexed by Web of Science.

[9] "Desk accepted" articles generally contain non-academic content, *e.g.* announcements by the editorial team, changes to peer review procedures or notifications about the death of an individual who had a close connection to the journal.

3 Gender gaps in time and length of peer review

3.1 Conceptual framework

In this section, we define four measures that capture gender differences in the length of peer review for eventually accepted papers: (i) the gender gap in the amount of time referees take to evaluate papers; (ii) the gender gap in the round-by-round probability of acceptance; (iii) the gender gap in total rounds of review; and (iv) the gender gap in the effort authors expend revising their manuscripts.

Equation (1) measures the gender gap in the amount of time referees take to evaluate papers that they do not reject (stage 1.2 in Figure 2):

$$\Delta^R = \mathbb{E}\Big[\text{time}^R_{ij_ft} - \text{time}^R_{ij_mt} \mid D_{j_l \max t} = \text{Accept}, R_{ij_lt} \neq \text{Reject}, \mathbf{Q}_{j_ft} = \mathbf{Q}_{j_mt}\Big], \tag{1}$$

where $\text{time}^R_{ij_lt}$ is the length of time referee i spends evaluating manuscript j_l in round t, $D_{j_l \max t}$ is the editor's decision at the end of the final round ($\max t$), R_{ij_lt} is referee i's round t recommendation, $\mathbf{Q}_{j_lt}$ represents a vector of j_l's round t qualities relevant to R_{ijt}—*e.g.*, novelty, rigour and readability—and $l \in \{m, f\}$ denotes male- and female-authored papers, respectively.

Equation (2) measures the gender gap in the round-by-round probability that referees recommend acceptance:

$$\Delta^{\text{accept}} = \mathbb{E}\Big[\text{accept}_{ij_ft} - \text{accept}_{ij_mt} \mid D_{j_l \max t} = \text{Accept}, \mathbf{Q}_{j_ft} = \mathbf{Q}_{j_mt}\Big], \tag{2}$$

where accept_{ij_lt} is an indicator variable equal to 1 if referee i recommends accepting manuscript j_l in round t.

Equation (3) measures the gender gap in rounds of review for eventually accepted papers:

$$\Delta^{\max t} = \mathbb{E}\big[\max t_{j_f} - \max t_{j_m} \mid D_{j_l \max t} = \text{Accept}, \mathbf{Q}_{j_f} = \mathbf{Q}_{j_m}\big], \tag{3}$$

where $\max t_{j_l}$ is manuscript j_l's total rounds of review and $\mathbf{Q}_j$ captures j's quality throughout the review process.

Finally, Equation (4) measures gender differences in the effort authors expend revising their papers (Figure 2, stage 0 for $t > 0$):

$$\Delta^A = \mathbb{E}\Big[\text{effort}^A_{j_ft} - \text{effort}^A_{j_mt} \mid D_{j_l \max t} = \text{Accept}, D_{j_lt-1} = \text{Revise}, \mathbf{Q}_{j_ft} = \mathbf{Q}_{j_mt}, \mathcal{I}_{j_f} = \mathcal{I}_{j_m}\Big], \tag{4}$$

where $\text{effort}^A_{j_lt}$ refers to effort spent revising manuscript j_l in round $t > 0$, D_{j_lt-1} is the editor's decision at the end of the previous round and $\mathcal{I}_{j_l}$ captures how accurately j_l's authors interpret and respond to R_{ijt}.

Δ^R, Δ^{accept}, $\Delta^{\max t}$ and Δ^A all condition on eventual acceptance and the selection criteria at *Energy Economics*. We discuss the former issue in Section 2.1. The latter assumes that female-authored papers are only compared to male-authored papers that have an equally legitimate claim to acceptance—either because they are, effectively, identical manuscripts, or (more likely) because their strengths and weaknesses evenly match.

Δ^R and Δ^A additionally condition on $R_{ij_lt} \neq$ Reject and $D_{j_lt-1} =$ Revise, respectively. The latter accounts for the fact that effort^A_{jt} is only possible to observe if revisions were requested in the previous round. The former restricts the population to referees who do not recommend rejection. There is a non-linear relationship between time^R_{ijt} and R_{ijt} (Appendix A.1)—*i.e.*, referees spend less time reviewing papers they reject compared to papers they don't, but more time on papers when recommending major instead of minor revisions or acceptance. They may also be slightly more likely to recommend rejecting

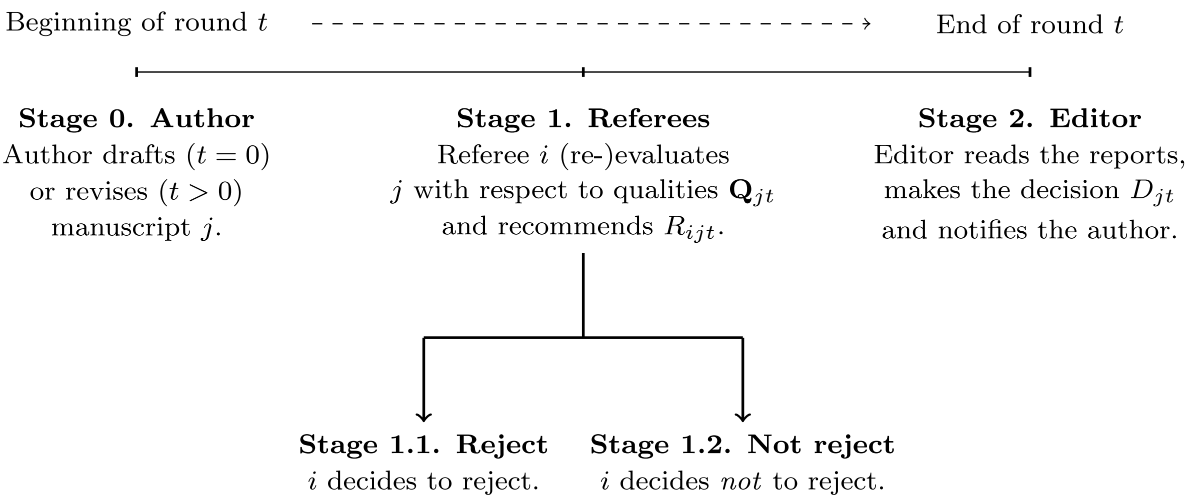

Figure 2: Sequence of events during revision round t

Note. 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 manuscript 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 referee i evaluates a vector of manuscript j's qualities $\mathbf{Q}_{jt}$ and summarises her opinion of them in the report and recommendation R_{ijt}. Round t concludes with the editor reading and processing the reports and recommendations from all referees 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 his manuscript, and round $t + 1$ begins.

female-authored papers relative to observably equivalent male-authored papers (Appendix A.3). As a result, aggregating time_{ijt} across all R_{ijt} may fail to accurately capture gender differences in the time referees spend reviewing manuscripts despite controlling for $\mathbf{Q}_{jt}$.

Finally, Δ^A conditions on authors being equally well-informed about the peer review process (*i.e.*, $\mathcal{I}_{j_f} = \mathcal{I}_{j_m}$). This rules out comparisons of effort between men and women who have different information about the appropriate way to interpret referee reports.

Throughout this paper, we interpret $\Delta^R = \Delta^{\text{accept}} = \Delta^{\max t} = \Delta^A = 0$ as desirable outcomes. Or in other words, we normatively assume that referees should not take longer to evaluate and accept male- and female-authored manuscripts—nor should female authors exert more effort revising those manuscripts—conditional on *Energy Economics*'s selection criteria. The validity of our interpretation depends on (i) the extent to which any non-zero gaps are counterbalanced by factors *outside* the narrow system we study; and (ii) how accurately we define, measure and condition on $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$. The latter issue is addressed in Section 3.2. We hope the present paper spurs further research on—and a healthy debate about—the former.

Conditional on these assumptions, we therefore interpret a positive Δ^R, $\Delta^{\max t}$, Δ^A and/or negative Δ^{accept} as some form of direct or systemic discrimination against female authors.[10] The first type of discrimination, 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 (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 care less about the impact a delayed and convoluted report can have on their careers.

[10] Our definitions of "direct" and "systemic" discrimination are similar to Bohren *et al.* (2022)'s: direct discrimination is differential treatment on the basis of a protected characteristic, holding other characteristics fixed; systemic discrimination is when outcomes are unnecessarily determined by non-gender characteristics that nevertheless disadvantage one gender over the other—*e.g.*, measuring employee performance using a proxy that ranks women lower than they would have been ranked had true performance been observed.

We define the second form of discrimination, statistical discrimination, to broadly include any kind of differential treatment that would not have occurred if information were perfect. Referees may spend more time evaluating—and subsequently ask for more changes to—female-authored manuscripts because they believe (correctly or incorrectly) that women submit lower quality work (Phelps 1972). They may also receive a signal about the quality of female-authored papers that is less precise (Aigner and Cain 1977)—perhaps because they find them harder to interpret or further away from their own expertise. And 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 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 (for a related discussion on race, see Small and Pager 2020). For example, if childcare responsibilities make it harder for women to attend conferences, conference proceedings may disproportionately publish manuscripts by male authors. Alternatively, if women are less likely to become influential in the profession, then referees may not have enough of an incentive to give them timely and coherent advice about how to revise their papers.

As most of these examples illustrate, each form of discrimination overlaps with and feeds back into other forms of discrimination—and imperfect information and power differentials are especially likely to combine in ways that disadvantage women. For example, 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 may 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. Similarly, widespread gender disparities (of any origin) probably spawn inaccurate stereotypes (Lang and Spitzer 2020; Reskin 2012). As a result, correlation between gender and organisational rank combined with the latter's use as a proxy for quality may cause some referees to mistakenly infer that gender *also* proxies for quality.

3.2 Estimation strategy

Δ^R, Δ^{accept} and $\Delta^{\max t}$ are ideally estimated using Equations (5), (6) and (7):

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

$$\text{accept}_{ijt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \, t + \tau_j + \epsilon_{ijt}, \tag{6}$$

and

$$\max t_j = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_j + \tau_j + \varepsilon_j, \tag{7}$$

where female_j is a binary variable equal to one if manuscript j is female-authored and accepted by referee i in round t, τ_j is a vector of year fixed effects, ε_{ijt} and ε_j are error terms, and time^R_{ijt}, accept_{ijt}, $\max t_j$, and quality vectors $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ are defined as in Section 3.1.

Applying Equations (5), (6) and (7) to the data poses two problems. First, each equation's β_1 consistently estimates Δ^R, Δ^{accept} and $\Delta^{\max t}$ only if referee assignment does not depend on author gender conditional on selection criteria.[11] For this reason, we account for year of submission in all three equations and round in Equations (5) and (6). In Appendix C, we additionally control for editor, secondary *JEL* code and referee fixed effects as well as institutional rank, author prominence, number of co-authors and manuscript

[11] In Equations (5) and (6), referee assignment cannot depend on author gender conditional on round; in Equation (7), they must be independent across all rounds.

length.

Our second problem is that we do not know the components included in the quality vectors $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ nor do we know how each is weighted in referees' and editors' overall decisions. We nevertheless attempt to (imperfectly) proxy for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ in various ways. Our primary proxy is citations. Yet, even if citations captured final acceptance criteria with zero non-classical measurement error—and evidence suggests that they do not—citations are neither round-specific nor measured pre-treatment, and thus may be influenced by peer review in ways that correlate with female_j.[12] For example, suppose referees work harder to improve women's manuscripts; then controlling for citations will exaggerate Δ^R, Δ^{accept} and $\Delta^{\max t}$. We therefore also proxy for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ using referees' and editors' round specific decisions (R_{ijt} and D_{jt}, Equation (5) only), the editor's decision in the previous round (D_{jt-1}, Equation (6) only), and the editor's decision in the first round (D_{j0}, Equation (7) only). Because each is undoubtedly affected by the same discrimination captured by Δ^R, Δ^{accept} and $\Delta^{\max t}$, controlling for them may cause us to under-estimate true gender gaps. Nevertheless, as long as at least one proxy is not biased in favour of women, then β_1 should still conservatively estimate Δ^R, Δ^{accept} and $\Delta^{\max t}$.

Estimating Δ^A poses an additional difficulty: we do not observe effort^A_{jt}. Instead, we proxy for it using the time authors spend revising their manuscripts each round:

$$\text{time}^A_{jt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\mathbf{Q}_{jt} + \beta_3 \, t + \beta_4 \,\mathcal{I}_j + \tau_j + \varepsilon_{jt}, \tag{8}$$

where time^A_{jt} is time spent revising manuscript j in round t and $\mathcal{I}_j$ is the maximum number of manuscripts previously published in *Energy Economics* by any author on j. As in Equations (5), (6) and (7), we control for year of submission and round; to proxy for $\mathbf{Q}_{jt}$, we use editors' decisions at the end of the previous round (D_{jt-1}), at the end of the current round (D_{jt}) and citations.

Unfortunately, time^A_{jt} is a poor proxy for effort^A_{jt}. As a result, β_1 in Equation (8) likely contains substantial non-classical measurement error. For example, greater childcare obligations, teaching loads or service responsibilities may force women to push their eventual resubmission dates further into the future, while fewer grants can leave 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 spend more time searching for and applying to jobs.[13] Because we have no reliable means of accounting for these factors, we encourage additional caution when interpreting Equation (8).

3.3 Results

3.3.1 *Energy Economics*

Table 1 displays results from OLS estimation of Equations (5) (columns (1)–(2)), (6) (column (3)), (7) (column (4)) and (8) (columns (4)–(5)). In the top panel, we proxy for gender using the gender of the corresponding author; in the bottom panel, we compare solo-female-authored papers to solo-male-authored papers. Results in both panels suggest that referees take longer reviewing and accepting female-authored papers, female-authored papers go through more rounds of review and their authors spend longer revising them.

First, referees spend, on average, 4–10 days longer reviewing female-authored papers. Columns (1) and (2) of Table 1 display results from OLS estimation of Equation (5). In column (1), we proxy for $\mathbf{Q}_{jt}$ using citations; our results suggest that referees take 4.5 days longer to review papers by female corresponding

[12] According to past research, 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 editors and referees give female-authored papers lower evaluations, conditional on citations.

[13] Although each of these examples could itself be interpreted as a form of discrimination, they occurred before peer review at *Energy Economics* and so are not captured by Δ^R, Δ^{accept}, Δ^{rounds} or Δ^A.

Table 1: Gender gaps in time and length of peer review at *Energy Economics*

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	4.482***	4.754***	−0.028**	0.103**	11.133**	10.890**
	(1.647)	(1.677)	(0.014)	(0.046)	(5.216)	(5.377)
t (round)	−15.893***	−12.608***	0.363***		−41.755***	−28.627***
	(0.972)	(1.625)	(0.016)		(2.340)	(2.182)
citations (asinh)	−5.243***		0.002	0.011	−11.604***	
	(0.836)		(0.004)	(0.015)	(1.796)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.413***				
		(1.798)				
revise (major)		7.187***				
		(2.280)				
D_{it} (editor's decision)						
revise (major)		2.963				2.382
		(3.215)				(7.482)
revise (minor)		2.643				5.581
		(1.986)				(4.495)
$\mathcal{I}_j$ (prominence)					−1.263***	−1.477***
					(0.367)	(0.278)
D_{jt-1} (major)						43.731***
						(4.204)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.083	0.071	0.275	0.038	0.114	0.136
Bounds (β_1)	[4.24, 4.48]	[4.75, 4.78]	[-0.04, -0.03]	[0.10, 0.10]	[9.64, 11.13]	[9.15, 10.89]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Solo-authored papers						
female	9.571**	8.852**	−0.013	0.071	30.741*	30.856*
	(3.975)	(4.104)	(0.035)	(0.122)	(17.150)	(17.670)
t (round)	−12.250***	−8.661***	0.329***		−34.898***	−27.559***
	(1.308)	(1.853)	(0.029)		(5.465)	(5.634)
citations (asinh)	−4.099***		0.002	−0.021	−5.132*	
	(1.277)		(0.009)	(0.040)	(2.697)	
R_{ijt} (referee's recommendation)						
revise (minor)		9.758**				
		(4.218)				
revise (major)		12.792***				
		(4.525)				
D_{it} (editor's decision)						
revise (major)		−0.801				−7.009
		(6.185)				(11.758)
revise (minor)		3.893				13.383
		(5.145)				(9.917)
$\mathcal{I}_j$ (prominence)					−3.901**	−3.528**
					(1.589)	(1.451)
D_{jt-1} (major)						27.690***
						(9.829)
No. obs.	1,275	1,275	1,350	447	744	744
R^2	0.069	0.067	0.259	0.047	0.102	0.117
Bounds (β_1)	[9.57, 10.20]	[8.76, 8.85]	[-0.03, -0.01]	[0.07, 0.07]	[26.01, 30.74]	[26.22, 30.86]
Mean dep. var.	53.395	53.395	0.245	1.669	67.227	67.227
Year	✓	✓	✓	✓	✓	✓

Note. Figures correspond to coefficients from estimating Equations (5), (6), (7) and (8). In the top panel, we proxy for gender using the gender of the corresponding author; in the bottom panel, we compare solo-female-authored papers to solo-male-authored papers. Standard errors clustered by referee (columns (1)–(2)), manuscript (column (3)) and corresponding author (columns (5)–(6)) in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

authors relative to papers by male corresponding authors, and 9.6 days longer to review papers by solo female authors compared to papers by male solo authors. Both gender gaps are roughly the same when we proxy for $\mathbf{Q}_{jt}$ using R_{ijt} and D_{jt} (column (2)).

Second, referees are less likely to recommend accepting female-authored papers. Column (3) in Table 1 shows results from estimating Equation (6) using a linear probability model. Every round of review, referees are about 3 percent less likely to recommend accepting papers by female corresponding authors compared to similar quality papers by male corresponding authors. When we restrict the sample to solo-authored papers, β_1 drops to −1 percent and its standard error rises. Although the sample size is much smaller, the gap is roughly similar when we proxy for $\mathbf{Q}_{jt}$ using the editor's decision in the previous round (for female corresponding authors, $\beta_1 = -0.036$, standard error 0.025).

Third, female-authored papers go through more rounds of review. Column (4) presents results from OLS estimation of Equation (7). Female-authored papers go through 0.07–0.1 more rounds compared to similar quality male-authored papers, although the gap is not significant for solo female-authored papers. When we account for $\mathbf{Q}_{jt}$ using editors' decisions in the first round, β_1 falls and is more noisily estimated (for female corresponding authors, $\beta_1 = 0.057$, standard error 0.042). Combined with results in column (3), this may suggest that women go through additional rounds because referees initially rate their papers more poorly relative to similar quality male-authored papers.[14]

Fourth, female authors take 11–31 days longer to revise their papers compared to men. Columns (5) and (6) display results from OLS estimation of Equation (8). Using citations again as a proxy of quality, column (5) shows that in each round of review, female corresponding authors take 11 days longer revising their papers and female solo authors take 31 days longer. Results are similar when accounting for $\mathbf{Q}_{jt}$ using editors' decisions (D_{it-1} and D_{it} in column (6)).

The gender gaps in Table 1 are robust to controlling for editor, secondary *JEL* code, referee and institution fixed effects as well as author prominence, number of co-authors and manuscript length (Appendix C); they also largely replicate using several alternative ways to capture a paper's gender composition (Appendix A.5). In Appendix A.3, we show additional evidence consistent with results in column (3)—namely, that referees reviewing female-authored papers are more likely to recommend major instead of minor revisions. Finally, to assess estimates' sensitivity 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 reports these bounds on β_1 corresponding to the assumption that the unobservables explain about as much of the variation in the dependent variables as the observables do.[15] In all instances, bounds form reasonably narrow neighbourhoods around point estimates of β_1.

Coefficients on the remaining variables in Table 1 mostly correspond to intuition. 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 and are more likely to recommend acceptance as t increases. Meanwhile, authors take longer when editors ask for major (instead of minor) revisions or when they have less experience with *Energy Economics*'s peer review process.[16] Interestingly, citations do not significantly correlate with probability of acceptance or rounds of review.

[14]Results are almost identical if we additionally control for citations (for female corresponding authors, $\beta_1 = 0.058$, standard error 0.042). Furthermore, conditional on recommending revisions, referees are more likely to suggest major instead of minor changes when evaluating female-authored papers (Appendix A.3).

[15]Specifically, we assume that the R^2 from a regression of the observables and unobservables on the dependent variable is no more than $2\,\widetilde{R}^2 - \mathring{R}^2$, where $\widetilde{R}^2$ is the R^2 from regressing the dependent variable on all observables and $\mathring{R}^2$ is the R^2 from only regressing the dependent variable on female_j.

[16]In column (2), the coefficients on "revise (major)" and "revise (minor)" are relative to the base level "accept"; in column (6), the coefficient on "D_{jt-1} (major)" compares time spent responding to major (instead of minor) revisions.

3.3.2 32 economics and finance journals

Table 2 displays results from estimating similar gender gaps in time and length of peer review for an additional 32 economics and finance journals (excluding *Energy Economics*). Unlike the comprehensive information we have for *Energy Economics*, the data are far more limited: they only include three dates (first submission, first decision and final acceptance) and do not contain referee recommendations or interim editorial decisions. Nevertheless, they suggest gender gaps in time and rounds of review that are roughly in line with those presented in Table 1.

To approximate time spent with referees, we take the difference (in days) between a paper's submission and first decision dates (*i.e.*, the date of the $t = 0$ editorial decision, see Figure 2).[17] Results from regressing this variable on proxies for a paper's gender composition and citations are presented in the first two columns of Table 2. In contrast with the results in Table 1, estimates in column (1) suggest that referees spend about the same amount of time evaluating female-authored papers, at least in the first round of review. As illustrated in Appendix B.3, however, these gaps vary widely across journals. Moreover, they negatively correlate with total rounds of review, suggesting that a true first round gap may be masked by female-authored papers that were subsequently subject to particularly tough peer review (Appendix B.4).[18] Column (2) corroborates this conclusion. It restricts the sample to papers that were accepted after a single round (the most likely outcome); the resulting estimates are roughly in line with those in Table 1.

In the third column of Table 2, we show results from regressing papers' total rounds of review on their gender composition and citations. Consistent with results in Table 1, female-authored manuscripts go through more rounds of review. The gap for female corresponding authors is smaller than the one estimated in Table 1; it is similar (and more precisely estimated) for solo-female-authored papers.

To approximate time spent with authors, we take the difference (in days) between a paper's first and final decision dates, excluding papers that were accepted in round $t = 0$ "as is" (*i.e.*, without any revisions). Because this figure includes both the time authors spend revising their papers as well as the time referees and editors subsequently take to review their revisions, we additionally control for a paper's total rounds of review (column (4)) and restrict the sample to manuscripts accepted after a single round (column (5)). Results in the top panel of column (4) are roughly in line with corresponding estimates in Table 1; results in the bottom panel are smaller in magnitude, but more precisely estimated. Compared to time spent with referees, there is also more homogeneity across journals—estimates shown in Appendix B.3 suggest a positive gap for the vast majority. In column (5), we restrict the sample to papers that were accepted after a single round; the gender gap roughly halves.

The gender gaps in Table 2 are robust to controlling for author prominence, number of co-authors and manuscript length (Appendix C). As in Table 1, bounds corresponding to the assumption that the unobservables explain about as much of the variation in the dependent variables as the observables do are generally tightly compacted around point estimates of female authorship.

Most coefficients on remaining variables are similar to corresponding estimates in Table 1. Highly cited papers are reviewed and revised faster and less prominent authors take more time revising; in contrast to results in Table 1, however, citations positively correlate with rounds of review.

[17] We drop 41 manuscripts with a first decision time exceeding 2.5 years. (Almost all were published in only two journals, and their delayed decisions likely reflect initially lost or forgotten submissions.)

[18] We observe a similar relationship in the data for *Energy Economics*.

Table 2: Gender gaps in time and length of peer review at 32 economics and finance journals

	Time to first decision		No. rounds ($\max t$)	Time from first to final decision	
	All papers	$\max t \leq 1$	All papers	All papers	$\max t = 1$
	(1)	(2)	(3)	(4)	(5)
Female corresponding authors					
female	0.431	3.566**	0.043**	14.562***	5.888**
	(0.442)	(1.611)	(0.020)	(3.039)	(2.336)
citations (asinh)	−5.018***	−4.490***	0.015*	−11.236***	−3.897**
	(0.996)	(1.422)	(0.008)	(2.680)	(1.531)
$\max t$				93.558***	
				(8.062)	
prominence				−0.835**	−0.876***
				(0.364)	(0.290)
No. obs.	24,511	10,708	24,560	23,907	10,083
R^2	0.166	0.210	0.228	0.408	0.247
Bounds (female)	[0.43, 3.51]	[3.57, 6.39]	[0.04, 0.04]	[11.20, 14.56]	[5.24, 5.89]
Mean dep. var.	130.491	127.557	1.779	214.954	112.311
Solo-authored papers					
female	0.290	8.436	0.078***	22.091***	12.409*
	(3.051)	(5.040)	(0.024)	(5.025)	(6.879)
citations (asinh)	−5.532***	−5.513***	0.029***	−11.767***	−4.535**
	(1.341)	(1.868)	(0.007)	(2.947)	(2.008)
$\max t$				95.992***	
				(10.333)	
prominence				−6.543***	−4.046***
				(1.693)	(1.052)
No. obs.	6,532	2,876	6,553	6,380	2,715
R^2	0.166	0.232	0.261	0.421	0.271
Bounds (female)	[0.29, 2.50]	[7.52, 8.44]	[0.07, 0.08]	[11.15, 22.09]	[10.64, 12.41]
Mean dep. var.	138.246	135.225	1.778	208.014	102.014
Journal	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓

Note. Top panel proxies for a paper's gender composition using the gender of its corresponding author; bottom panel compares solo-female-authored papers to solo-male-authored papers. The dependent variables are the number of days between a paper's submission and first decision dates (columns (1)–(2)), total rounds of review (column (3)) and the number of days between a paper's first and final decision dates (columns (4)–(5)). Column (2) includes only papers that were accepted without revisions or after a single round ($t \leq 1$); column (5) excludes papers that were revised multiple times ($t = 1$). Standard errors clustered by journal and authors' countries in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

4 The role of statistical discrimination

Our results thus far suggest that referees take longer reviewing and accepting female-authored papers ($\Delta^R > 0$ and $\Delta^{\text{accept}} < 0$, respectively), manuscripts by women go through more rounds of review ($\Delta^{\max t} > 0$) and their authors spend longer revising them (consistent with, although not definitive evidence of, $\Delta^A > 0$). As discussed in Section 3.1, we interpret these findings as some form of direct or systemic taste-based discrimination, disparate treatment resulting from power differentials or statistical discrimination, where we broadly define the latter to include any kind of differential treatment that would not have occurred if information were perfect.

In this section, we focus on better understanding the particular contribution of direct or systemic statistical discrimination. Our own subjective belief is that most people in the profession do not harbour strong, prejudicial preferences against female economists or their research. Although we are more convinced that power differentials play a role, they are arguably more difficult to pro-actively address in the short to medium term. Furthermore, any power or taste differentials that *do* exist are probably exacerbated by imperfect information; thus, by studying the latter dimension, we may indirectly shed light on how to alleviate discrimination on the former two dimensions as well.

4.1 Conceptual framework

To understand how information (or a lack thereof) can drive gender gaps in time and length of peer review, suppose referees had perfect information when reviewing for *Energy Economics*. Such knowledgeable referees 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. Perfectly informed referees should also have no trouble clearly specifying the changes that would need to be made before a paper could be published. Thus, in a world of complete and perfect information, positive Δ^R, $\Delta^{\max t}$, Δ^A and negative Δ^{accept} are independent of statistical discrimination, and instead indicate differential treatment motivated purely by taste and/or power differentials.

Of course, information is neither complete and perfect nor costless to obtain, but by a similar logic, better informed referees should be better at distinguishing between low- and high-quality papers compared to poorly informed referees, all else equal. They are also undoubtedly more knowledgeable and confident about the standards of acceptance at *Energy Economics*. Consequently, the signals they receive about the quality of submitted papers are probably more accurate—and contain fewer gender-specific distortions—than the signals received by poorly informed referees. Furthermore, signal clarity likely gives rise to more precise and easier-to-implement reports, all else equal. Thus, if the magnitudes of Δ^R, $\Delta^{\max t}$, Δ^A and Δ^{accept} decline as information improves (holding everything else constant), then statistical discrimination was an important factor driving their original non-zero values.

To operationalise this insight, define $f_j(i)$ as the function that assigns referee i to paper j and let Equation (9) represent the change in an outcome Δ^s with respect to information:

$$\tilde{\Delta}^s = \mathbb{E}\left[\Delta_p^s - \Delta_w^s \mid f_j(i_p) = f_j(i_w)\right], \tag{9}$$

where $s \in \{R, \text{accept}, \max t, A\}$, Δ_p^s and Δ_w^s measure Δ^s when referee i is poorly (p) and well-informed (w), respectively, and $f_j(i_p) = f_j(i_w)$ assumes that both versions of i review paper j.

$\tilde{\Delta}^s$ conditions on the same factors as Δ^s, *i.e.*, non-rejection, the selection criteria at *Energy Economics* as well as $R_{ijt} \neq \text{Reject}$ ($\tilde{\Delta}^R$ only), $D_{jt-1} = \text{Revise}$ ($\tilde{\Delta}^A$ only) and $\mathcal{I}_{j_f} = \mathcal{I}_{j_m}$ ($\tilde{\Delta}^A$ only). It also adds a crucial new assumption, namely that poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the same manuscripts.

To specifically illustrate how Equation (9) separates factors related to incomplete information from those that are independent of it, suppose $\Delta_p^R > 0$ and $\Delta_w^R < \Delta_p^R$. This pattern suggests that poorly informed referees statistically discriminate against female authors and, as a result, spend longer reviewing their papers; as information improves, however, their ability to identify quality in men's and women's papers converges, so Δ_w^R declines relative to Δ_p^R. On the other hand, if $\Delta_w^R \geq \Delta_p^R$, then the time referees spend evaluating female-authored papers is independent of (or even negatively related to) information. This outcome suggests that the gender gap is not driven by statistical discrimination.

4.2 Estimation strategy

To identify the effect of information, we re-estimate the equations in Section 3.2 including an indicator variable of referee information and its interaction with female. Thus, β_1 in Equations (10), (11), (12) and (13) estimates Δ_w^R, Δ_w^{accept}, Δ_w^{rounds} and Δ_w^A, respectively, $\beta_1 + \beta_3$ estimates Δ_p^R, Δ_p^{accept}, Δ_p^{rounds} and Δ_p^A, and β_3 estimates $\tilde{\Delta}^R$, $\tilde{\Delta}^{\text{accept}}$, $\tilde{\Delta}^{\text{rounds}}$ and $\tilde{\Delta}^A$:

$$\text{time}_{ijt}^R = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\text{info}_{ij}^p + \beta_3 \,\text{female}_j \times \text{info}_{ij}^p + \beta_4 \,\mathbf{Q}_{jt} + \beta_5 \,t + \tau_j + \varepsilon_{ijt}, \tag{10}$$

$$\text{accept}_{ijt} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\text{info}_{ij}^p + \beta_3 \,\text{female}_j \times \text{info}_{ij}^p + \beta_4 \,\mathbf{Q}_{jt} + \beta_5 \,t + \tau_j + \varepsilon_{ijt}, \tag{11}$$

$$\max t_{ij} = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\text{info}_{ij}^p + \beta_3 \,\text{female}_j \times \text{info}_{ij}^w + \beta_4 \,\mathbf{Q}_j + \tau_j + \varepsilon_{ij}, \tag{12}$$

$$\text{time}_{ijt}^A = \beta_0 + \beta_1 \,\text{female}_j + \beta_2 \,\text{info}_{jt}^p + \beta_3 \,\text{female}_j \times \text{info}_{jt}^p + \beta_4 \,\mathbf{Q}_{jt} + \beta_5 \,t + \beta_6 \,\mathcal{I}_j + \tau_j + \varepsilon_{ijt}, \tag{13}$$

where time_{ijt}^A is time spent revising paper j in round t given j was refereed by i, $\max t_{ij}$ is j's total rounds of review conditional on i and info_{ij}^p is an indicator variable equal to one if i was poorly informed. Remaining variables are defined in Section 3.2. As before, we proxy for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ using citations, and $\mathcal{I}_j$ using the maximum number of manuscripts previously published in *Energy Economics* by any author on j.

Although "poorly informed" referees are hard to identify, we attempt to do so using their experience reviewing for *Energy Economics*. In particular, we set $\text{info}_{ij}^p = 1$ if referee i had only reviewed four or fewer eventually accepted papers when (s)he was asked to review paper j and 0 otherwise.[19] Accounting for identical assumptions to those discussed in Section 3.2—and assuming that novice referees are indeed more poorly informed about the peer review process than expert referees, all else equal—β_1 in Equations (10), (11) and (12) is an unbiased estimate of Δ_w^R, Δ_w^{accept} and Δ_w^{rounds}, respectively and $\beta_1 + \beta_3$ is an unbiased estimate of Δ_p^R, Δ_p^{accept} and Δ_p^{rounds}. By further assuming that poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the exact same manuscripts, β_3 is also an unbiased estimate of $\tilde{\Delta}^R$, $\tilde{\Delta}^{\text{accept}}$ and $\tilde{\Delta}^{\text{rounds}}$.

As discussed previously, time_{ijt}^A remains a problematic proxy for effort in Equation (13). Nevertheless, $\beta_1 \leq 0$ plausibly implies that $\beta_1 + \beta_3$ conservatively estimates Δ_p^A, and β_3 conservatively estimates $\tilde{\Delta}^A$. To see why, let $k \in \{p, w\}$ capture referee i_k's information (poorly and well-informed, respectively), define $\text{time}_{jt}^A(i_k)$ as the amount of time j's authors spend responding to i_k and suppose that, conditional on selection criteria and authors' information, gender differences in time_{jt}^A are a linear function of effort, referee information and discrimination:

$$\text{time}_{j_f t}^A(i_k) - \text{time}_{j_m t}^A(i_k) = \eta \left(\text{effort}_{jt}^A + \text{effort}_{jt}^A(i_p) \times \text{info}_{ij}^p\right) \text{female}_j + \left(\theta + \zeta \,\text{info}_{ij}^p\right) \text{female}_j, \tag{14}$$

where info_j^p is defined in Equation (13), $\text{effort}_{jt}^A \geq 0$ is effort spent responding to i_w, $\text{effort}_{jt}^A(i_p)$ is the additional effort required to respond to i_p, θ and $\theta + \zeta$ correspond to Δ_k^A when $k = w$ and $k = p$, respectively, ζ specifically captures statistical discrimination by poorly informed referees (*i.e.*, $\tilde{\Delta}^A$), and

[19]See Appendix A.4 for further discussion and descriptive statistics on this variable.

η absorbs the extra time women spend completing their revisions, holding effort fixed.[20]

β_1 in Equation (13) is the marginal change in time_{ijt}^A with respect to author gender when papers are assigned well-informed referees. It corresponds to Equation (14) when $k = w$ (and thus $\text{info}_{ij}^p = 0$):

$$\beta_1 \equiv \Delta_w^A + \eta \, \text{effort}_{jt}^A. \tag{15}$$

Similarly, $\beta_1 + \beta_3$ in Equation (13) is the marginal change in time_{ijt}^A with respect to gender when papers are assigned to poorly informed referees:

$$\beta_1 + \beta_3 \equiv \Delta_p^A + \eta \left(\text{effort}_{jt}^A + \text{effort}_{jt}^A(i_p) \right). \tag{16}$$

Finally, β_3 in Equation (13) is the marginal effect of gender when switching from a well-informed referee to a poorly informed one. It is simply the difference between Equations (15) and (16):

$$\beta_3 \equiv \tilde{\Delta}^A + \eta \, \text{effort}_{jt}^A(i_p). \tag{17}$$

Equations (15), (16) and (17) make clear that β_1, $\beta_1 + \beta_3$ and β_3 reflect discrimination *and* the non-classical measurement error introduced by women potentially spending longer completing their revisions, conditional on effort.

Assume that well-informed referees are not biased *in favour* of women—*i.e.*, $\Delta_w^A \geq 0$. Since $\text{effort}_{jt}^A \geq 0$ (by definition), $\beta_1 \leq 0$ therefore implies that $\eta \leq 0$ (Equation (15)). Or in other words, female-authored papers are revised just as quickly as male-authored papers, conditional on effort.[21] Furthermore, $\text{effort}_{jt}^A(i_p) \geq 0$ follows from our assumptions that experience is a valid proxy for referee information and poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the exact same manuscripts. Thus $\eta \leq 0$ implies that $\beta_1 + \beta_3 \leq \Delta_p^A$ (Equation (16)) and $\beta_3 \leq \tilde{\Delta}^A$ (Equation (17)). Moreover, by explicitly ruling out positive discrimination by well-informed referees, β_3 specifically captures the cost to women (in terms of extra days spent revising) of statistical discrimination by poorly informed referees.

4.3 Results

Evidence in Figure 3 consistently suggests that poorly informed referees presumably statistically discriminate against female authors. It plots β_1, $\beta_1 + \beta_3$ and β_3 from estimating Equations (10), (11), (12) and (13) with data from *Energy Economics*; left-hand-side graphs proxy for female_j using the gender of the corresponding author; graphs on the right-hand-side compare solo-female-authored papers to papers by male corresponding authors.[22] Novice referees have reviewed four or fewer eventually accepted papers; experts have reviewed five or more.

Consider first solo-female-authored papers. On average, novice referees spend 6.2 more days reviewing solo-female-authored papers compared to papers by male corresponding authors (standard error 3.9), or 6.5 days longer after accounting for referee fixed effects (standard error 5.9); however, gender gaps for experienced referees are –2.7 (standard error 9.9) and –2.1 (standard error 4.4), respectively. Thus,

[20] effort_{jt}^A and $\text{effort}_{jt}^A(i_p)$ capture authors' total effort in the absence of discrimination; ζ picks up whatever time and effort poorly informed referees *additionally* impose on female authors, including, *e.g.*, the extra time women need to manage childcare so that they can exert the added effort required to address statistical discrimination.

[21] $\beta_1 \leq 0$ is not necessarily evidence that women have fewer time constraints compared to men. For example, women may be more likely to prioritise finishing an R&R over other tasks (possibly because they face stricter tenure requirements or are employed in more precarious positions).

[22] We exclude all instances of the editor-in-chief acting as a referee as many of these interventions are made for administrative reasons, do not reflect meaningful peer review and would otherwise dominate our sample of expert referees. Because very few referees have reviewed solo-authored papers by both genders, we also compare solo female-authored papers to papers by male corresponding authors, in contrast to results shown in the bottom panel of Table 1.

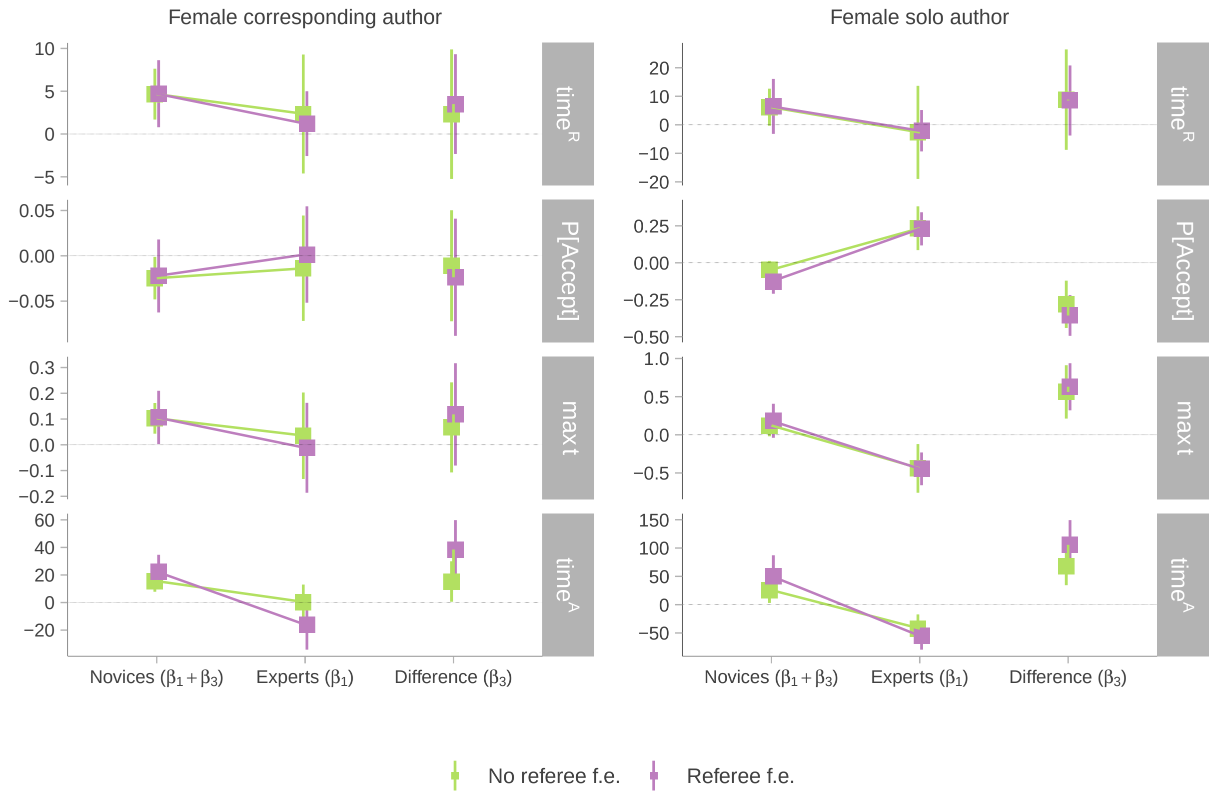

Figure 3: The relationship between referee experience and length of peer review

Note. Graph displays β_1, $\beta_1 + \beta_3$ and β_3 from estimating Equations (10), (11), (12) and (13) with the data from *Energy Economics*. Left-hand-side graphs proxy for female$_j$ using the gender of the corresponding author; graphs on the right-hand-side compare solo-female-authored papers to papers by male corresponding authors. Novice referees have reviewed four or fewer eventually accepted papers; experts have reviewed five or more. Lines represent 90 percent confidence internals; standard errors clustered at the referee level.

gender gaps in time spent with referees are 8.8 days larger among novice referees than they are among expert referees, and 8.5 days larger conditional on referee fixed effects. (However, standard errors on both differences are large: 10.7 and 7.5, respectively.)

Each round, novice referees are 5 percent less likely to recommend accepting papers by female solo authors compared to papers by male corresponding authors (standard error 0.04), and 13 percent less likely after accounting for referee fixed effects (standard error 0.05). On the other hand, expert referees are 23 percent *more* likely to accept solo-female-authored papers with and without referee fixed effects (standard errors 0.09 and 0.07, respectively). Gender gaps in probability of acceptance are therefore 28 percentage points higher (in absolute value) among novice referees than they are among expert referees (standard error 0.10), and 36 percentage points higher after accounting for referee fixed effects (standard error 0.08).

Similarly, when evaluated by novice referees, solo-female-authored papers go through 0.12 more rounds of review compared to papers by male corresponding authors (standard error 0.09)—and 0.18 more rounds conditional on referee fixed effects (standard error 0.14). But again, both gaps fall among experienced referees (–0.44 (standard error 0.19) and –0.45 (standard error 0.13), respectively). As a result, gender gaps are 0.56 rounds larger when papers are reviewed by novices instead of experts (standard error 0.21), and 0.63 rounds larger conditional on referee (standard error 0.19).

Finally, solo female authors reviewed by novice referees spend, on average, 26 more days (per round) revising their papers compared to male corresponding authors (standard error 14), and 51 more days after accounting for referee fixed effects (standard error 22). When reviewed by expert referees, however, female authors spend 42 *fewer* days revising compared to men (standard error 15), or 55 fewer days after accounting for referee fixed effects (standard error 14.8). Moreover, under plausible assumptions, $\beta_1 \leq 0$ implies that female-authored papers are revised just as quickly as male-authored papers, holding effort fixed (see Section 4.2); as a result, our estimates of β_3 in Figure 3 suggest that statistical discrimination by novice referees costs women at least 68 extra revision days each round (standard error 20)—and 106 additional days conditional on referee fixed effects (standard error 27).

Results for papers by female corresponding authors are analogous (albeit weaker): estimates of $\beta_1 + \beta_3$ are almost always large and statistically significantly different from zero, whereas estimates of β_1 are not. Their differences, however, are noisy—β_3 is significant at traditional thresholds only when estimating Equation (13).

Combined, evidence in Figure 3 points to statistical discrimination by novice referees: the magnitude of every gender gap is greater when papers are evaluated by poorly informed referees. When reviewed by well-informed referees, however, all four gender gaps disappear and possibly reverse. As we show in Appendix C, these conclusions are robust to controlling for editor, secondary *JEL* code, institutional rank fixed effects as well as author prominence, number of co-authors and manuscript length; they also largely replicate using several alternative ways to capture a paper's gender composition (Appendix A.5).

Nevertheless, our conclusions *do* depend on several assumptions. First, β_1 and $\beta_1 + \beta_3$ are unbiased (or conservative) estimates of Δ_p^s and Δ_w^s ($s \in \{R, \text{accept}, \text{rounds}, A\}$) only if citations are not biased in favour of women conditional on *Energy Economics*'s selection criteria. Although it is impossible to entirely rule out the converse, we find similar results when proxying for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ using referees' and editors' round specific decisions (Equations (10)), editors' decisions in the previous round (Equation (11)), editors' decisions in the first round (Equation (12)) and editors' round-specific decisions (Equation (13)) (see Appendix A.6). Furthermore, a large body of research consistently finds that citations are, if anything, biased in favour of men (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Ferber 1988; Koffi 2019).

Second, interpreting β_3 as an unbiased estimate of $\tilde{\Delta}^s$ requires that poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the exact same manuscripts. We explore this assumption more carefully in Appendix A.4, but do not find consistent evidence that it does not hold, particularly after controlling for referee fixed effects. Nevertheless, we encourage additional caution when interpreting these findings, given referees may have been non-randomly assigned in ways we did not anticipate.

Third—and most controversially—$\beta_1 + \beta_3$ and β_3 in Equation (13) are unbiased estimates of Δ_p^A and $\tilde{\Delta}^A$ only if well-informed referees are not biased in favour of women, conditional on the selection criteria at *Energy Economics*. Although β_1 is never distinguishable from zero in Equation (13), it is significantly negative and positive for female solo authors in Equations (11) and (12), respectively. These results are consistent with bias in favour of women but could also be evidence that our proxies for quality under-estimate the value of female-authored papers in ways that well-informed referees do not.[23] Because we cannot differentiate between these two possibilities, we also encourage additional caution when interpreting $\beta_1 + \beta_3$ and β_3 in Equation (13).

[23]Evidence in Hengel and Moon (2022) and Zacchia (2021) suggests that women's contributions to research are (unintentionally) discounted in many popular proxies of academic impact. Consistent with these conclusions, we find that papers by female solo authors published in *Energy Economics* and the 32 additional journals are cited more than papers by male solo authors (Appendix E). (Evidence is somewhat more mixed when gender is determined by the gender of the corresponding author.)

5 Conclusion

In this paper, we describe and analyse gender differences in the length of peer review. Using detailed administrative data from an economics field journal, we find that in each round of review, referees spend 4–10 more days reviewing papers by female authors; they are also 1–3 percent less likely to recommend accepting them. Compared to similar quality male-authored papers, female-authored papers go through 0.07–0.1 more rounds of review and their authors spend 11–31 more days revising them each round. Less disaggregated data from an additional 32 economics and finance journals largely corroborate these results.

These gender gaps add up. The cumulative time gap is 29–60 days at *Energy Economics* and 18–29 days at the other 32 economics and finance journals (Appendix D). A back-of-the-envelope calculation suggests that for every 10 papers a woman publishes in these journals, she will have spent 6–20 months 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 2022), we believe our estimates may be lower bounds on the gendered time cost of peer review.

Furthermore, we find that all gender gaps are greater when papers are reviewed by novice referees than when they are reviewed by expert referees, even after conditioning on referee fixed effects. 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 gender gaps in the length of peer review decline (and possibly reverse).

Unfortunately, our data cannot precisely identify why novice referees statistically discriminate. There is no strong direct connection to field, institutional rank or author prominence (see Appendix C), but many other possibilities remain: for example, novice referees may have biased beliefs about female authors; alternatively, smaller networks—possibly exacerbated by authors suggesting their own referees—may mean referees are simply less familiar with women's work (see, *e.g.*, Ductor *et al.* 2021).

One possible policy response is to limit the amount of time referees and authors are given to review and revise their papers. However, this response does not address the poor quality information likely driving these gaps; as a consequence, such a policy could easily backfire in various ways. For example, if poorly informed referees are given less time to evaluate papers, they might end up submitting lower quality reports; similarly, forcing authors to return their revisions more quickly might handicap women who still have to convincingly respond to more sceptical referees.

Instead, we propose two alternative policy solutions. The first disproportionately assigns female-authored papers to more experienced referees. This policy directly addresses the disproportionate burden referees' imperfect information clearly has on female authors. Importantly, it also does not disrupt the communication channels between author and referee nor does it obviously decrease the informativeness of the refereeing process.

We also believe our results call for simultaneously expanding 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 believe an interesting avenue for future research could be to empirically test the efficacy of these programmes.[24]

[24] In this sense, we also contribute to a small—but growing—empirical literature studying how to develop the skills required to be an effective peer reviewer. While numerous articles provide excellent advice on how to peer review manuscripts (see, *e.g.*, Berk *et al.* 2015; Kelly *et al.* 2014), there is very little empirical research exploring the effectiveness of this advice. A notable exception is Davis *et al.* (2020), who use interview and survey data to identify and study the skills most relevant to funding peer review. One of their conclusions is that reviewers often obtain relevant skills by observing the behaviours and actions of more experienced reviewers and participating more frequently in grant review panels. Consistent with these

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 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

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.

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.

Angus, S. D. *et al.* (2021). "Geographic diversity in economic publishing". *Journal of Economic Behaviour and Organization* 190, pp. 255–262.

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.

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.

Bohren, J. A., P. Hull, and A. Imas (2022). "Systemic discrimination: theory and measurement". Mimeo.

Bowles, H. R., L. McGinn Kathleen, and L. Babcock (2005). "Constraints and triggers: situational mechanics of gender in negotiation". *Journal of Personality and Social Psychology* 89 (6), pp. 951–965.

Brock, J. M. and R. De Haas (2022). "Discriminatory lending: evidence from bankers in the lab". *American Economic Journal: Applied Economics* 15 (2), pp. 31–68.

Bucher-Koenen, T., A. Hackethal, and J. Koenen (2021). "Gender differences in financial advice". Mimeo.

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

Coffman, K. B., M. R. Collis, and L. Kulkarni (2023). "Whether to apply". *Management Science* (forthcoming).

Davis, M. L. E. S. *et al.* (2020). "What makes an effective grants peer reviewer? An exploratory study of the necessary skills". *PLOS ONE* 15 (5), e0232327.

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

Delfino, A. (2022). "Breaking gender barriers: experimental evidence on men in pink-collar jobs". Mimeo.

results, we find that experience in the peer review process and familiarity with the standards of acceptance at a particular journal also help referees achieve more equitable outcomes.

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 (2021). “Gender and collaboration”. *Review of Economics and Statistics* (forthcoming).

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”. *Nature Neuroscience* 23 (8), pp. 918–926.

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

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

Gallen, Y. and M. Wasserman (2022). “Informed choices: gender gaps in career advice”. Mimeo.

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.

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

Hengel, E. and E. Moon (2022). “Gender and quality at top economics journals”. Mimeo.

Kelly, J., T. Sadeghieh, and K. Adeli (2014). “Peer review in scientific publications: benefits, critiques, and a survival guide”. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine* 25 (3), pp. 227–243.

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.

Matthews, S. A., M. Okuno-Fujiwara, and A. Postlewaite (1991). “Refining cheap-talk equilibria”. *Journal of Economic Theory* 55 (2), pp. 247–273.

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.

RePEc (2023). *Female representation in Economics, as of March 2023.* Date accesssed: 2023-04-29. URL: https://ideas.repec.org/top/female.html.

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

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

Appendices

A *Energy Economics*, supplemental material

A.1 Non-monotonic relationship between R_{ijt} and time_{ijt}^{R}

As we discuss in Section 3.1, one reason we restrict $R_{ijt} \neq$ Reject in Equation (1) 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 for this is presented in Figure A.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.

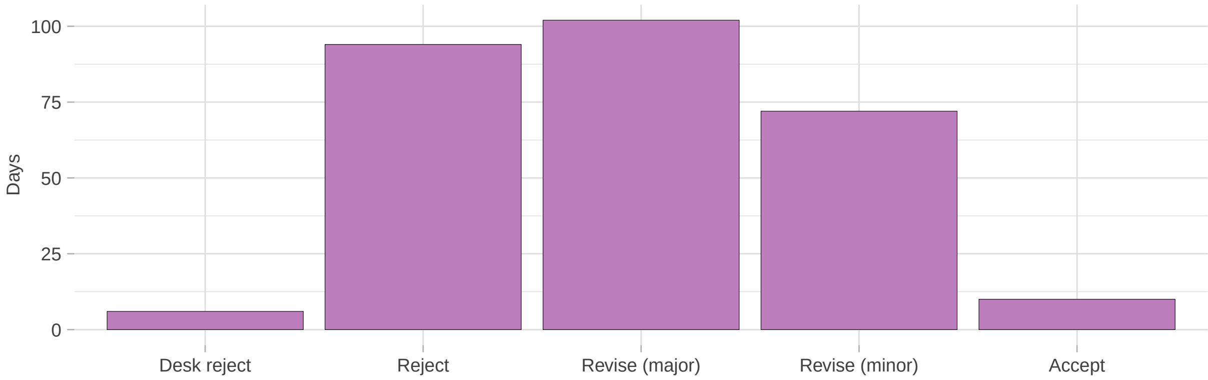

Figure A.1: Median number of days spent with referees and editors, by editor decision

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

The sample used to estimate Figure A.1 includes 13,396 unique manuscript-round instances, of which 3,774 were desk rejected manuscripts, 3,437 were manuscripts rejected after being sent out for review, 2,052 were "revise (major)" decisions, 1,767 were revise (minor) decisions and 2,366 were accept decisions. Note that these 2,366 accepted observations correspond to the same unique 2,359 full-length, regular issue manuscripts analysed in the main body of the paper. (Seven manuscripts went through an additional round of review after being accepted.)

A.2 Round-by-round descriptive statistics

Table A.1: Round-by-round descriptive statistics for papers published in *Energy Economics*

		Percentage of papers						
		Female		Decision			Avg. no. days with	
Round	Obs.	Corr.	Solo	Major rev.	Minor rev.	Accept	Refs.+Eds.	Authors
0	2,359	14.5	3.6	72.4	26.7	0.9	152.5	
1	2,337	14.6	3.7	11.9	36.5	51.6	57.9	97.4
2	1,136	16.9	4.3	4.4	19.6	76.0	22.5	32.6
3	274	15.7	2.6	4.0	16.4	79.6	12.9	22.5
4+	79	15.2	7.9	6.3	20.3	73.4	7.9	16.7

Note. Table displays round-specific descriptive statistics for papers published in *Energy Economics*. The first column is is the round, the second column is the number of observations and the third and fourth columns are the percentages of manuscripts with a female corresponding and solo author, respectively. Columns 5–7 are the percentages 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 spends each round with referees/editors and with authors.

In Table A.1, we summarise basic statistics on the manuscripts in our *Energy Economics* sample. In the initial round of review (round 0), there are 2,359 manuscripts in our data, of which 15 percent have a female corresponding author and 4 percent were solo-female-authored. 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.)

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 female-authored papers 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+.

A.3 Round-by-round outcome probabilities

In Table A.2 we report results from regressing an indicator variable equal to 1 if a referee recommended rejection (columns (1) and (3)) or recommended major revisions conditional on not recommending rejection (columns (2) and (4)) on author gender, round, citations (asinh) and year fixed effects. In the first panel, "female" refers to papers with a female corresponding author; in the second panel, the sample is restricted to solo-authored papers.

Table A.2: Round-by-round probabilities at *Energy Economics*

	Female corresponding author		Female solo-authored	
	$\mathbb{P}$[Reject]	$\mathbb{P}$[Revise (major)]	$\mathbb{P}$[Reject]	$\mathbb{P}$[Revise (major)]
	(1)	(2)	(3)	(4)
female	0.013	0.047***	−0.021	0.013
	(0.012)	(0.016)	(0.025)	(0.040)
t (round)	0.002	−0.262***	0.001	−0.261***
	(0.007)	(0.012)	(0.015)	(0.026)
citations (asinh)	0.001	−0.010*	−0.014*	−0.013
	(0.004)	(0.006)	(0.008)	(0.011)
No. obs.	5,439	7,044	1,019	1,275
R^2	0.004	0.127	0.011	0.136
Bounds (β_1)	[0.01, 0.01]	[0.05, 0.05]	[-0.02, -0.02]	[0.01, 0.02]
Mean dep. var.	0.079	0.344	0.074	0.369
Year	✓	✓	✓	✓

Note. Figures correspond to coefficients from regressing an indicator variable equal to 1 if a referee recommended rejection (columns (1) and (3)) or major revisions conditional on not recommending rejection (columns (2) and (4)) on author gender, round, citations (asinh) and year fixed effects. Standard errors clustered by referee in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Referees are about 1 percent more likely to recommend rejecting papers with a female corresponding author compared to similar quality papers by male corresponding authors, although the estimate is very noisily estimated (column (1)). Results in column (2) are more precise, and suggest that, conditional on being accepted or offered the opportunity to revise, papers by female corresponding authors are about 5 percent more likely to be asked to make major changes.

In the second panel of Table A.2 we restrict the sample to solo-authored papers. We find female solo-authored papers are about 2 percent less likely to be rejected compared to similar quality solo male-authored papers; however, the effect is noisily estimated—indeed, there are only 8 instances (from 7 unique manuscripts) of referees recommending rejection for solo female-authored papers. In column (4) we find that referees who do not recommend rejection are about 1 percent more likely to recommend female-authored papers undergo major revisions, although, again, the effect is very noisily estimated.

A.4 Referee experience

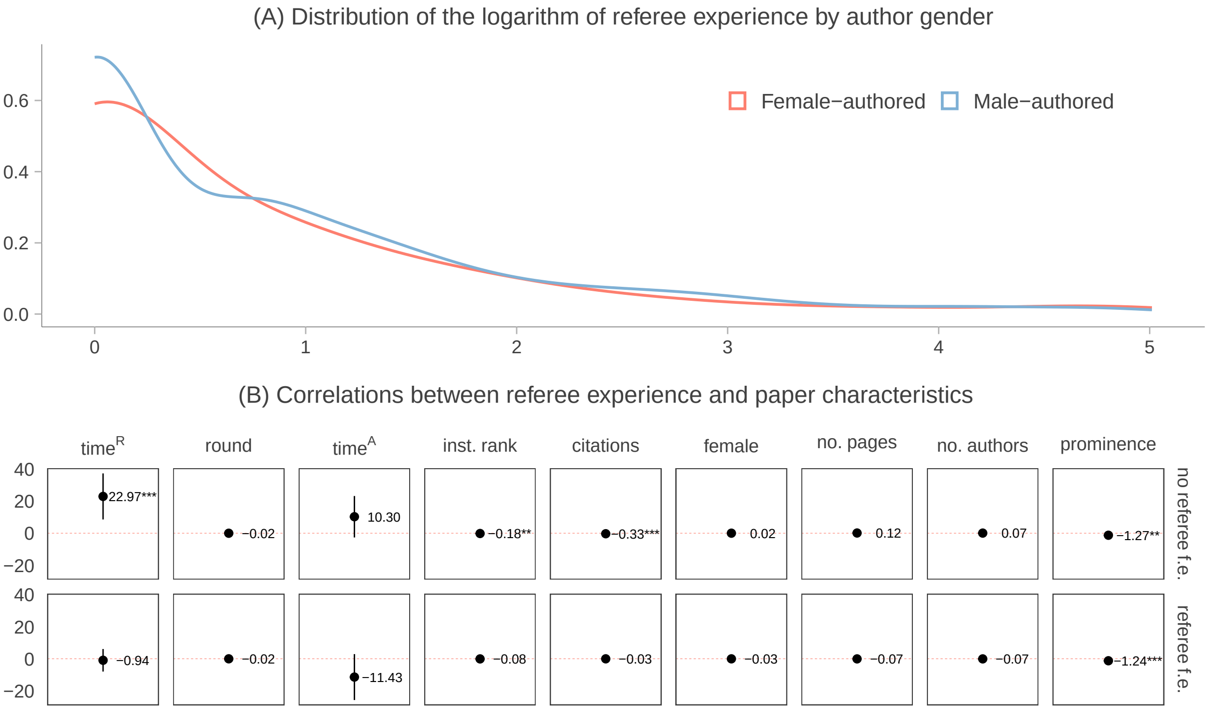

Figure A.2: The relationship between referee experience and paper characteristics

Note. Top 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 an indicator variable equal to 1 for novice referees (*i.e.*, referees who previously reviewed four or fewer eventually accepted papers). Regressions control for submission year fixed effects and citations (where relevant); standard errors clustered by referee. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Of the 4,724 unique referee-manuscript instances in our data, 82 percent are of "novice" referees (*i.e.*, referees who have previously reviewed 1–4 eventually accepted manuscripts, including the current paper) and 18 percent are "expert" referees (*i.e.*, referees who have previously reviewed 5 or more eventually accepted manuscripts, including the current paper). Among papers with a female corresponding author, 83 percent are reviewed by novice referees; among papers with a male corresponding author, 81 percent are; the difference (0.02) is not statistically significant at traditional thresholds (standard error 0.02).

Figure A.2A plots the distribution of referee experience. Referees with no prior experience reviewing for *Energy Economics* are slightly more likely to review male-authored papers; referees who had already reviewed 1–2 papers are slightly more likely to review female-authored papers. Otherwise, men's and women's papers appear to be reviewed by similarly experienced referees.

As discussed in Section 4.1, identifying $\tilde{\Delta}^s$, $s \in \{R, \text{accept}, \max t, A\}$ requires that poorly informed referees' evaluations of manuscripts are compared to well-informed—but otherwise identical—referees' evaluations of the exact same manuscripts. This rules out situations where poorly informed referees are assigned manuscripts that are harder to evaluate for reasons that are independent of quality. For example, suppose poorly and well-informed referees are asked to review the same manuscript, but well-informed referees are assigned the English version, whereas poorly informed referees are assigned its Latin translation. Assuming most referees are more fluent in English than they are in Latin, poorly informed referees would therefore probably have a harder time evaluating the paper compared to well-informed referees, all else equal.

It is impossible to fully assess the validity of this assumption. Nevertheless, we initially attempt to do so in Figure A.2B. It plots coefficients from nine separate OLS regressions of paper characteristics on

Table A.3: Correlations between referee experience, paper characteristics and author gender

	citations (asinh)		author prominence		institutional rank	
	(1)	(2)	(3)	(4)	(5)	(6)
female corr.	−0.210**	−0.032	−2.219***	−1.017**	0.057	0.327**
	(0.105)	(0.091)	(0.746)	(0.401)	(0.168)	(0.154)
novice ref.	−0.347***	−0.036	−1.453**	−1.338***	−0.140	−0.011
	(0.120)	(0.076)	(0.633)	(0.353)	(0.100)	(0.114)
female×novice ref.	0.142	0.066	1.488**	0.720	−0.301*	−0.603***
	(0.115)	(0.121)	(0.745)	(0.460)	(0.178)	(0.183)
citations (asinh)			0.847***	0.521***	0.230***	0.211***
			(0.104)	(0.086)	(0.022)	(0.031)
No. obs.	4,724	4,724	4,724	4,724	4,724	4,724
R^2	0.448	0.777	0.176	0.688	0.042	0.582
Mean dep. var.	2.700	2.700	3.299	3.299	2.880	2.880
female solo	−0.370	−0.090	−2.834**	−1.642	0.019	0.332
	(0.544)	(0.520)	(1.321)	(1.074)	(0.809)	(0.764)
novice ref.	−0.322***	−0.053	−1.375***	−1.028***	−0.097	0.028
	(0.111)	(0.092)	(0.531)	(0.351)	(0.096)	(0.148)
female×novice ref.	0.295	0.075	1.272	0.368	−0.359	−0.496
	(0.560)	(0.560)	(1.337)	(1.169)	(0.823)	(0.802)
citations (asinh)			0.814***	0.507***	0.202***	0.230***
			(0.091)	(0.117)	(0.024)	(0.036)
No. obs.	3,293	3,293	3,293	3,293	3,293	3,293
R^2	0.426	0.795	0.193	0.732	0.036	0.620
Mean dep. var.	2.714	2.714	3.282	3.282	2.834	2.834
Year	✓	✓	✓	✓	✓	✓
Referee		✓		✓		✓

Note. Table displays coefficients from regressing citations (asinh), author prominence and institutional rank on citations (where appropriate), female authorship, a dummy variable for novice referees, and an interaction between the latter two variables. Standard errors clustered on referees in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

a dummy variable for novice referees, controlling for submission year fixed effects and citations (where relevant). Without referee fixed effects (top row), less experienced referees may be assigned papers that take longer to review, accumulate fewer citations and are by less prominent authors at lower ranked institutions (remaining estimates are not significantly different from zero). After controlling for referee fixed effects, coefficients are indistinguishable from zero with one exceptions: compared to expert referees, novice referees may be assigned manuscripts by less prominent authors.

Although Figure A.2B does not consistently suggest that novice referees are assigned more difficult to review papers compared to expert referees, editors could plausibly assign novice referees easier-to-evaluate male-authored papers and harder-to-evaluate female-authored papers and experienced referees the reverse in a way that washes out when aggregating over genders. We therefore additionally regress citations (asinh), author prominence and institutional rank on female-authorship, a dummy variable for poorly informed referees, their interaction and, where appropriate, citations. (See Appendix C.4 for details on how institutional rank was calculated.) Results are presented in Table A.3. Only estimates in columns (5) and (6) plausibly refute the validity of our assumption: they suggest that female-authored papers assigned to novice referees are harder to evaluate—as proxied for by institutional rank—relative to male-authored papers; the opposite is true for expert referees. (However, controlling for institutional rank in Figure 3 does not change our general conclusions, see Figure C.3, Appendix C.4.) In contrast, results in columns (1)–(4) suggest that female-authored papers assigned to novice referees are easier to evaluate compared to male-authored papers and the reverse for expert referees.

A.5 Alternative proxies for gender

The following tables and figures replicate Table 1 and Figure 3 using alternative definitions of female authorship.[1] Table A.4 and Figure A.3 compare entirely male-authored papers to entirely female-authored papers. Table A.5 and Figure A.4 compare papers with a senior female author to papers with a senior male author (mixed-gendered papers with a senior male author are excluded).[2] Table A.6 and Figure A.5 replace the binary variable of female authorship with a continuous measure of the ratio of female authors. In Table A.7 and Figure A.6 we define female authorship as having at least 50 percent female authors (mixed-gendered papers with fewer than 50 percent female authors are excluded). Finally, Table A.8 and Figure A.7 compare papers with at least one female author to exclusively male-authored papers.

In general, results in Tables A.4–A.8 are similar to those presented in Table 1. Results in columns (5)–(6) are most similar, although the gaps are larger in Tables A.4 and A.6 than they are when female authorship is defined less restrictively (*e.g.*, as in Table A.8). Results in columns (3) and (4) are same signed compared to corresponding estimates in Table 1, although they are more noisily estimated; in Tables A.7 and A.8, they are also much smaller in magnitude. In Tables A.4–A.6, gender gaps in days spent with referees (columns (1)–(2)) are similarly sized or larger compared to estimates in Table 1. The gaps are much smaller (and not significantly different from zero) in Tables A.7 and A.8.

Figures A.3–A.7 all suggest that time spent with authors decreases and probability of acceptance increases when papers are reviewed by experts instead of novices, in line with corresponding estimates shown in Figure 3. Figures A.3 and A.5 suggest that time spent with referees also declines, although the evidence in Figures A.4, A.6 and A.7 is less clear. In contrast with Figure 3, total rounds of review does not appear to decline when papers are reviewed by more experienced referees, suggesting that this effect may only be relevant to papers that are solo authored by a women or by a female corresponding author.

[1] We only have data on the gender of corresponding authors for the 32 economics and finance journals analysed in Table 2; we therefore do not replicate it using alternative proxies for author gender.

[2] Senior authorship is defined as having at least as many papers previously published in *Energy Economics* at the time of submission as any other co-author.

A.5.1 Exclusively female-authored

Table A.4: Table 1, exclusively female-authored

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
100% female	6.063**	6.443**	−0.026	0.093	22.393**	22.575*
	(2.990)	(3.066)	(0.024)	(0.084)	(11.278)	(11.897)
t (round)	−15.851***	−11.727***	0.358***		−40.615***	−27.838***
	(1.024)	(1.600)	(0.018)		(2.766)	(2.598)
citations (asinh)	−5.229***		0.007	0.000	−12.177***	
	(0.910)		(0.005)	(0.018)	(2.146)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.583***				
		(2.049)				
revise (major)		8.532***				
		(2.512)				
D_{it} (editor's decision)						
revise (major)		4.625				7.800
		(3.353)				(8.913)
revise (minor)		4.227*				7.472
		(2.334)				(5.284)
$\mathcal{I}_j$ (prominence)					−1.162***	−1.522***
					(0.384)	(0.309)
D_{jt-1} (major)						41.746***
						(5.073)
No. obs.	4,998	4,998	5,306	1,688	2,728	2,728
R^2	0.080	0.070	0.273	0.037	0.112	0.131
Bounds (β_1)	[4.96, 6.06]	[5.72, 6.44]	[-0.03, -0.03]	[0.09, 0.09]	[17.07, 22.39]	[17.44, 22.57]
Mean dep. var.	53.505	53.505	0.272	1.619	71.397	71.397
Year	✓	✓	✓	✓	✓	✓

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

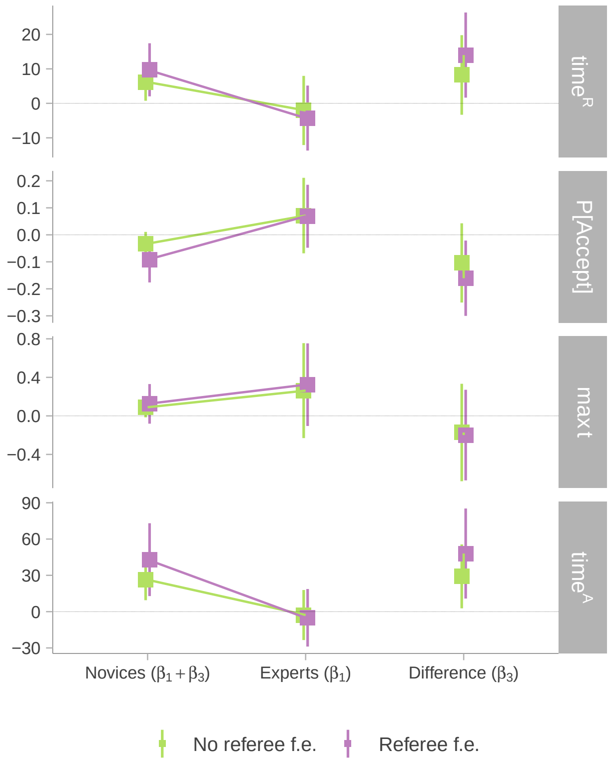

Figure A.3: Figure 3, exclusively female-authored

Note. Estimates are identical to those in Figure 3 except that the independent variable female has been replaced with a dummy variable equal to 1 if a paper is entirely female-authored and 0 if it is entirely male-authored. (Mixed-gendered papers are excluded.)

A.5.2 Senior female author

Table A.5: Table 1, senior female author

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
senior fem.	3.932***	4.386***	−0.019*	0.074*	8.299*	8.244*
	(1.424)	(1.465)	(0.011)	(0.040)	(4.784)	(4.829)
t (round)	−15.884***	−12.596***	0.363***		−41.731***	−28.562***
	(0.977)	(1.636)	(0.016)		(2.338)	(2.179)
citations (asinh)	−5.221***		0.002	0.011	−11.683***	
	(0.836)		(0.004)	(0.015)	(1.799)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.458***				
		(1.791)				
revise (major)		7.171***				
		(2.272)				
D_{it} (editor's decision)						
revise (major)		2.946				2.397
		(3.217)				(7.497)
revise (minor)		2.571				5.601
		(1.991)				(4.509)
$\mathcal{I}_j$ (prominence)					−1.181***	−1.398***
					(0.381)	(0.291)
D_{jt-1} (major)						43.873***
						(4.212)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.083	0.071	0.275	0.037	0.114	0.136
Bounds (β_1)	[3.17, 3.93]	[4.08, 4.39]	[-0.02, -0.02]	[0.07, 0.07]	[6.13, 8.30]	[6.01, 8.24]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Year	✓	✓	✓	✓	✓	✓

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

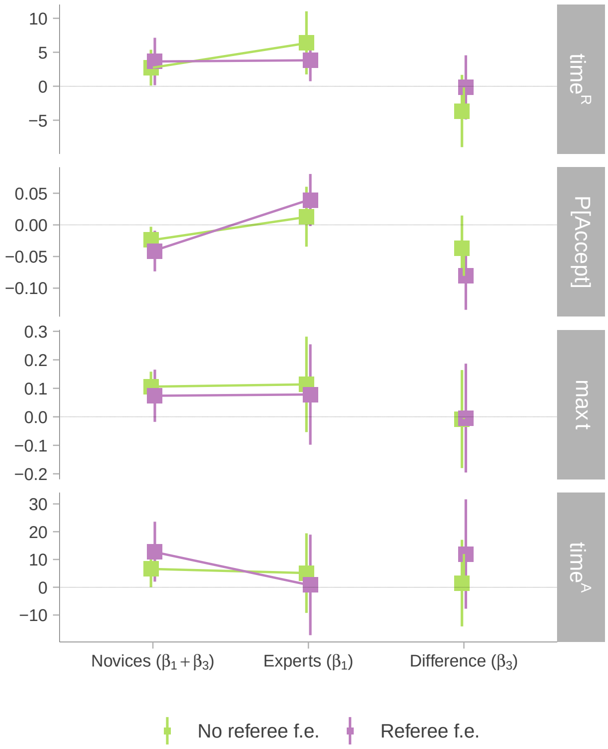

Figure A.4: Figure 3, senior female-author

Note. Estimates are identical to those in Figure 3 except that the independent variable female has been replaced with a dummy variable equal to 1 if a paper had at least one female author who had previously published at least as many papers in *Energy Economics* as her co-authors at the time the paper in question was published. (Mixed-gendered papers with a senior male co-author are excluded.)

A.5.3 Ratio of female authors

Table A.6: Table 1, ratio of female authors

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
fem. ratio	2.769	2.822	−0.018	0.074	15.227**	15.625*
	(2.187)	(2.249)	(0.019)	(0.062)	(7.761)	(7.976)
t (round)	−15.796***	−12.415***	0.362***		−41.613***	−28.514***
	(0.982)	(1.640)	(0.016)		(2.341)	(2.179)
citations (asinh)	−5.306***		0.002	0.010	−11.703***	
	(0.838)		(0.004)	(0.015)	(1.804)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.443***				
		(1.805)				
revise (major)		7.344***				
		(2.287)				
D_{it} (editor's decision)						
revise (major)		3.149				2.499
		(3.231)				(7.498)
revise (minor)		2.760				5.636
		(2.003)				(4.518)
$\mathcal{I}_j$ (prominence)					−1.338***	−1.543***
					(0.381)	(0.287)
D_{jt-1} (major)						43.812***
						(4.224)
No. obs.	7,006	7,006	7,432	2,346	3,798	3,798
R^2	0.081	0.068	0.274	0.037	0.114	0.137
Bounds (β_1)	[2.44, 2.77]	[2.54, 2.82]	[-0.02, -0.02]	[0.07, 0.07]	[13.35, 15.23]	[14.15, 15.62]
Mean dep. var.	53.359	53.359	0.272	1.622	71.231	71.231
Year	✓	✓	✓	✓	✓	✓

Note. Estimates are identical to those in Table 1 except that the independent variable female has been replaced with a continuous variable equal to the ratio of female authors on a paper. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

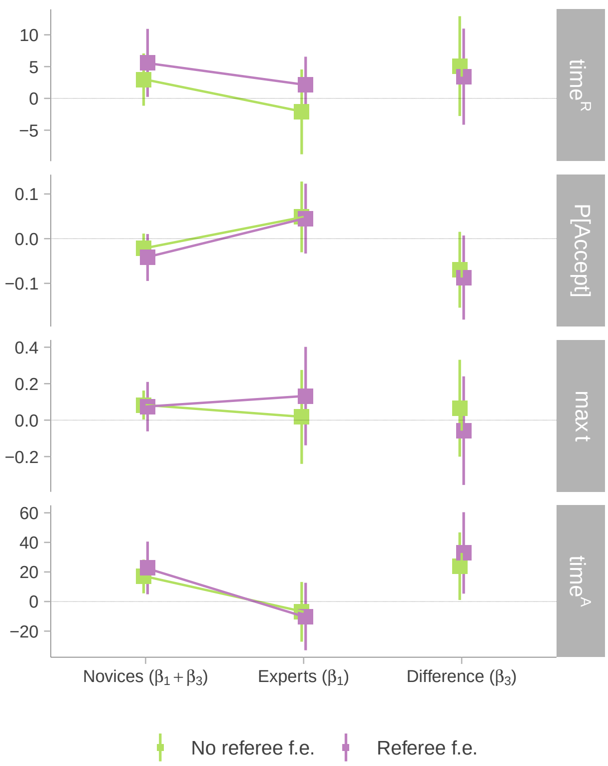

Figure A.5: Figure 3, ratio of female authors

Note. Estimates are identical to those in Figure 3 except that the independent variable female has been replaced with a continuous variable equal to the ratio of female authors on a paper.

A.5.4 50 percent female authors

Table A.7: Table 1, 50 percent female authors

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
50% female	0.481	0.430	−0.011	0.040	7.080	7.354
	(1.644)	(1.697)	(0.014)	(0.043)	(5.121)	(5.139)
t (round)	−15.713***	−11.650***	0.359***		−41.004***	−27.997***
	(0.976)	(1.523)	(0.017)		(2.528)	(2.354)
citations (asinh)	−5.392***		0.006	−0.002	−12.076***	
	(0.892)		(0.005)	(0.016)	(2.011)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.133***				
		(1.955)				
revise (major)		8.043***				
		(2.403)				
D_{it} (editor's decision)						
revise (major)		4.583				7.759
		(3.099)				(8.214)
revise (minor)		3.617*				8.084*
		(2.096)				(4.797)
$\mathcal{I}_j$ (prominence)					−1.512***	−1.791***
					(0.427)	(0.305)
D_{jt-1} (major)						42.416***
						(4.510)
No. obs.	5,924	5,924	6,297	1,997	3,236	3,236
R^2	0.078	0.066	0.273	0.036	0.113	0.134
Bounds (β_1)	[0.15, 0.48]	[0.05, 0.43]	[-0.01, -0.01]	[0.03, 0.04]	[5.93, 7.08]	[6.48, 7.35]
Mean dep. var.	53.266	53.266	0.271	1.623	71.437	71.437
Year	✓	✓	✓	✓	✓	✓

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

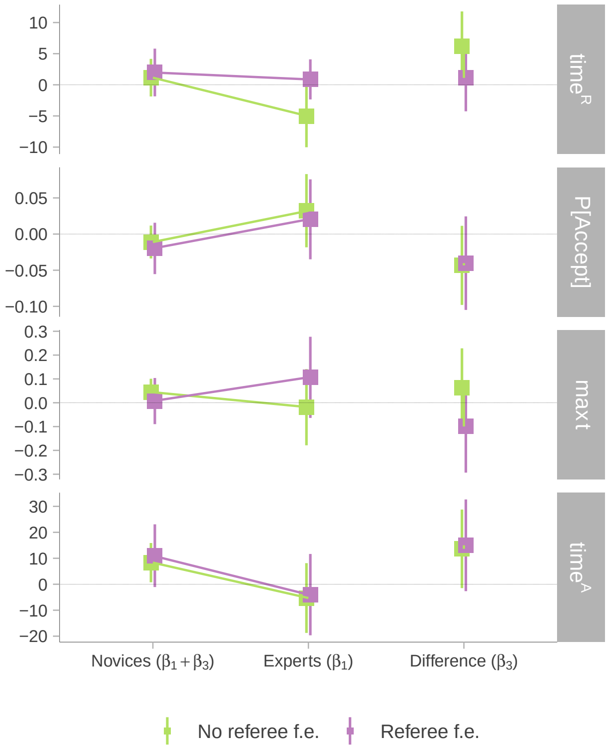

Figure A.6: Figure 3, 50 percent female authors

Note. Estimates are identical to those in Figure 3 except that the independent variable female has been replaced with a dummy variable equal to 1 if at least 50 percent of authors are female and 0 otherwise. (Mixed-gendered papers with fewer than 50 percent female authors are excluded.)

A.5.5 At least one female author

Table A.8: Table 1, at least one female author

	Eqn. (5) Days spent with referees		Eqn. (6) ℙ[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
1+ female	1.338	1.031	−0.006	0.028	5.689	5.694
	(1.178)	(1.205)	(0.010)	(0.034)	(3.871)	(3.911)
t (round)	−15.790***	−12.401***	0.362***		−41.603***	−28.488***
	(0.982)	(1.640)	(0.016)		(2.340)	(2.180)
citations (asinh)	−5.324***		0.002	0.009	−11.773***	
	(0.837)		(0.004)	(0.015)	(1.803)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.446***				
		(1.804)				
revise (major)		7.365***				
		(2.285)				
D_{it} (editor's decision)						
revise (major)		3.161				2.532
		(3.234)				(7.507)
revise (minor)		2.763				5.650
		(2.007)				(4.537)
$\mathcal{I}_j$ (prominence)					−1.379***	−1.589***
					(0.370)	(0.280)
D_{jt-1} (major)						43.867***
						(4.248)
No. obs.	7,006	7,006	7,432	2,346	3,798	3,798
R^2	0.081	0.068	0.274	0.036	0.113	0.136
Bounds (β_1)	[1.34, 1.88]	[1.03, 1.27]	[-0.01, -0.01]	[0.03, 0.03]	[5.69, 6.75]	[5.69, 6.76]
Mean dep. var.	53.359	53.359	0.272	1.622	71.231	71.231
Year	✓	✓	✓	✓	✓	✓

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

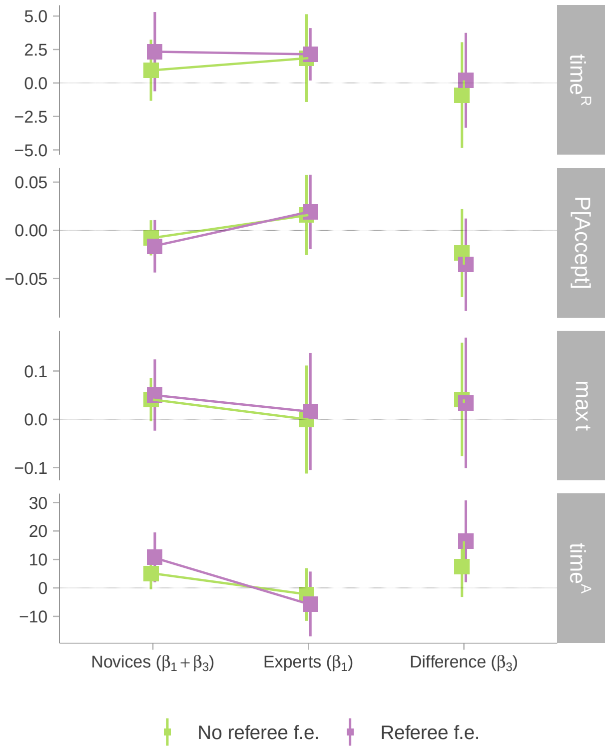

Figure A.7: Figure 3, at least one female author

Note. Estimates are identical to those in Figure 3 except that the independent variable female has been replaced with a dummy variable equal to 1 if at least one author on a paper is female.

A.6 Figure 3, accounting for alternative proxies of quality

As discussed in Section 4.3, β_1 and $\beta_1 + \beta_3$ in Figure 3 are unbiased (or conservative) estimates of Δ_p^s and Δ_w^s ($s \in \{R, \text{accept}, \text{rounds}, A\}$) only if citations are not biased in favour of women conditional on *Energy Economics*'s selection criteria. Although, a large body of research consistently finds that citations are, if anything, biased in favour of men (Dion *et al.* 2018; Dworkin *et al.* 2020; Ferber 1986; Ferber 1988; Koffi 2019), they are neither round-specific nor measured pre-treatment, and thus may be influenced by peer review in ways that correlated with female_j (for further discussion, see Section 3.2).

We therefore additionally proxy for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ using referees' and editors' round-specific decisions (Equation (10)), previous round decisions (Equation (11)), first round decisions (Equation (12)) and round-specific decisions (Equation (13)).[3] Results are shown in Figure A.8; they are very similar to those in Figure 3.

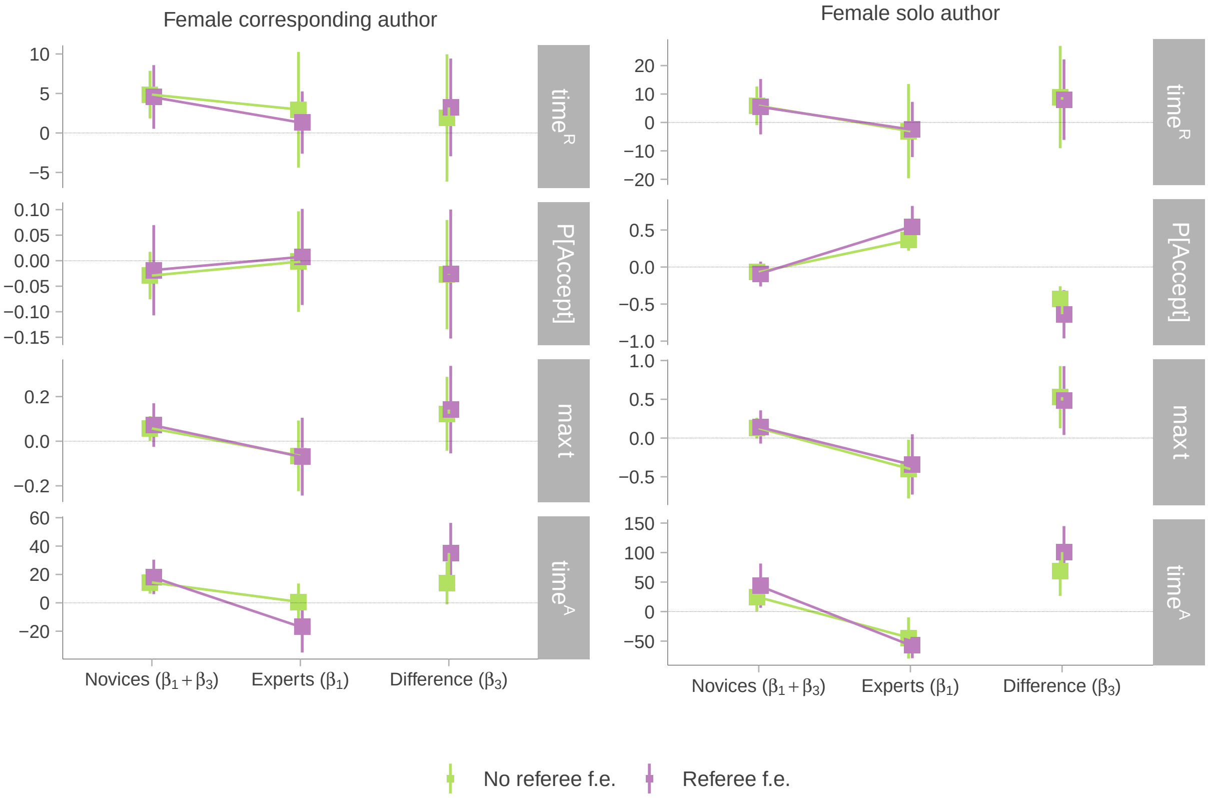

Figure A.8: Figure 3, accounting for alternative proxies of quality

Note. Figure is identical to Figure 3 except that we replace citations (asinh) as our proxy for $\mathbf{Q}_{jt}$ and $\mathbf{Q}_j$ with the following alternative proxies: referees' and editors' round-specific decisions (Equation (10)), editors' round-specific decisions (Equation (13)), editors' decisions in the previous round (Equation (11)) and editors' decisions in the first round (Equation (12)).

[3] In order to maximise the sample size—particularly when including referee fixed effects—we re-estimate Equation (13) without including editors' decisions in the previous round. (This contrasts with results shown in column (6) of Table 1 which includes this variable as an alternative proxy of quality.)

B 32 economics and finance journals, supplemental material

B.1 List of journals

Table B.1 lists the 32 economics and finance journals we describe in Section 2.2 and analyse in Section 3.3.2. Because journals began managing their submissions through EEM at different times, our sample is not balanced across journal and year.

Table B.1: The 32 economics and finance journals analysed in Section 3.3.2

	Years covered		
Journal	Submission year	Publication year	Obs.
Economic Modelling	2005–2019	2006–2019	2,919
Economic Systems	2008–2019	2009–2019	302
Economics Letters	2004–2019	2004–2019	4,675
European Economic Review	2002–2019	2003–2019	1,207
European Journal of Political Economy	2005–2018	2006–2019	618
Global Finance Journal	2014–2019	2015–2019	98
International Review of Economics and Finance	2010–2019	2010–2019	859
Japan and the World Economy	2005–2019	2005–2019	300
Journal of Banking and Finance	2007–2019	2008–2019	2,256
Journal of Behavioral and Experimental Finance	2013–2019	2014–2019	115
Journal of Commodity Markets	2015–2019	–	57
Journal of Corporate Finance	2004–2019	2005–2019	799
Journal of Development Economics	2004–2018	2005–2019	970
Journal of Economic Theory	2013–2019	2013–2019	546
Journal of Economics and Business	2006–2018	2007–2019	273
Journal of Empirical Finance	2005–2018	2006–2019	602
Journal of Environmental Economics and Management	2010–2019	2011–2019	448
Journal of Financial Intermediation	2013–2018	2013–2019	109
Journal of Financial Stability	2006–2018	2006–2019	434
Journal of Housing Economics	2006–2018	2006–2019	239
Journal of International Economics	2007–2018	2007–2019	786
Journal of Macroeconomics	2005–2019	2006–2019	774
Journal of Mathematical Economics	2005–2019	2006–2019	744
Journal of Monetary Economics	2004–2019	2005–2019	671
Journal of Multinational Financial Management	2005–2019	2005–2019	199
Journal of the Japanese and International Economies	2004–2019	2005–2019	294
North American Journal of Economics and Finance	2007–2019	2007–2019	587
Pacific-Basin Finance Journal	2006–2019	2007–2019	542
Regional Science and Urban Economics	2007–2019	2008–2019	627
Research in International Business and Finance	2006–2019	2006–2019	425
Resource and Energy Economics	2005–2018	2007–2019	376
Resources Policy	2007–2019	2008–2019	709

Note. First column lists the journals described and analysed in Sections 2.2 and 3.3.2; in the second and third columns are years covered by submission and publication year; in the final column are observation counts.

B.2 Summary statistics

In Figure B.1 we show summary statistics similar to those reported in Figure 1 but for the additional 32 economics and finance journals described in Section 2.2 and analysed in Section 3.3.2. We find trends very similar to those reported in Section 2.1.1.

As in *Energy Economics*, the percentage of papers with a female corresponding author hovers around 16 percent (Figure B.1A), papers with a female corresponding author go through more rounds of review (Figure B.1C), the number of manuscripts each journal publishes per year has been steadily increasing (Figure B.1B), while the average number of rounds per paper has stayed relatively flat (Figure B.1D). In contrast to data from *Energy Economics*, however, there has been a more noticeable increase in eventually published female-authored papers that were submitted between 2005–2017: in 2005, 14 percent of papers were female-authored; in 2017, 20 percent were.

Unlike our data for *Energy Economics*, we do not precisely know how much time authors spend revising their papers. To approximate it, we take the difference (in days) between a paper's first and final decision dates. Although this figure includes any time referees and editors subsequently take to review authors' revisions, 44 percent of papers are revised no more than once, so it should primarily reflect time spent with authors. Consistent with evidence from *Energy Economics*, this proxy does not change very much over our sample period (Figure B.1H); it also suggests that women spend more time revising compared to men (204 vs. 225 days, respectively; Figure B.1G).

Finally, Figures B.1E and B.1F plot the distribution and two-year moving averages of the difference (in days) between a paper's submission and first decision dates, which is meant to (imperfectly) proxy for time spent with referees. In contrast to data from *Energy Economics*, the average number of days with referees has been steadily declining since 2005. And while referees take slightly more time to review female-authored papers conditional on spending less than five months with a paper, they spend slightly less time with them if they take five months or more.

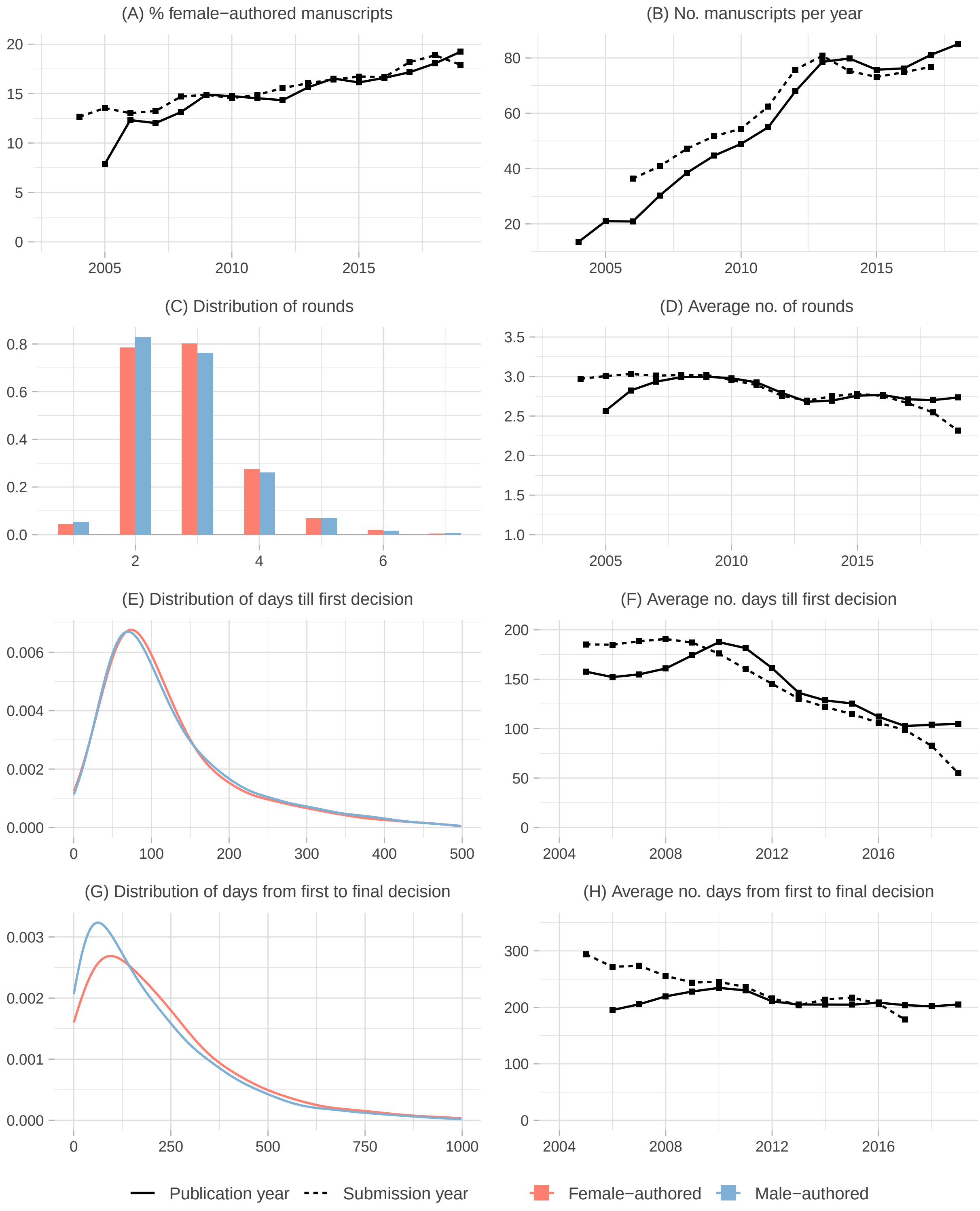

Figure B.1: Summary statistics on manuscripts published in 32 economics and finance journals

Note. Graph (A) displays the two-year moving average of the percentage of papers published and submitted each year with a female corresponding author; Graph (B) is the two-year moving average of the total number of manuscripts published and submitted each year. Graphs (C) and (D) plot the distribution of rounds by corresponding author gender and the number of rounds per year (two-year moving average), respectively. Graphs (E) and (F) plot the distribution of days from submission to first decision by corresponding author gender and the same number of days per year (two-year moving average), respectively. Graphs (G) and (H) show the distribution of days from first to final decision by corresponding author gender and the same number of days per year (two-year moving average), respectively.

B.3 Gender gap heterogeneity across journals

Figures B.2A–E plot the coefficient on female from regressions identical to those in the top panel of Table 2 except that samples are restricted to observations from particular journals; Figure B.2F similarly plots the gender gap in total time under review from the same regression used to estimate column (1) of Table D.2 (Appendix D). The coefficient on female when the sample is restricted to *Energy Economics* is highlighted in red on each graph;[4] for the reasons discussed in Section 2.2, Figure B.2 does not identify the other 32 journals.

There is quite a bit of heterogeneity across journals in gender gaps in time and length of peer review, and particularly in the time to first decision when the sample includes all papers (Figure B.2A): some journals have a significant gender gap favouring men; for others, the gap significantly favours women. When the sample is restricted to papers that were either accepted without revision or after a single round ($t \leq 1$) the gender gap generally favours men, although it is rarely significant at traditional thresholds (Figure B.2B).

For 80 percent of journals, the time from first to final decision is wider for women than it is for men (Figure B.2C). However, the gap is much smaller conditional on a paper only undergoing a single revision ($t = 1$) and is negative in about a third of journals (Figure B.2D).

In the final row of Figure B.2 we plot the coefficient on female when the dependent variable is total rounds of review (Figure B.2E) and the total time from first submission to final acceptance (Figure B.2F). Although papers with a female corresponding author go through more rounds of review, for a non-negligible number of them, the reverse is true; however, total time spent in review is almost always longer for women and never significantly below zero at any journal.

[4] To generate comparable estimates for *Energy Economics* in Figure B.2, we applied the sample restrictions and data definitions described in Section 3.3.2. As a result, these estimates are not directly comparable to those shown in Table 1.

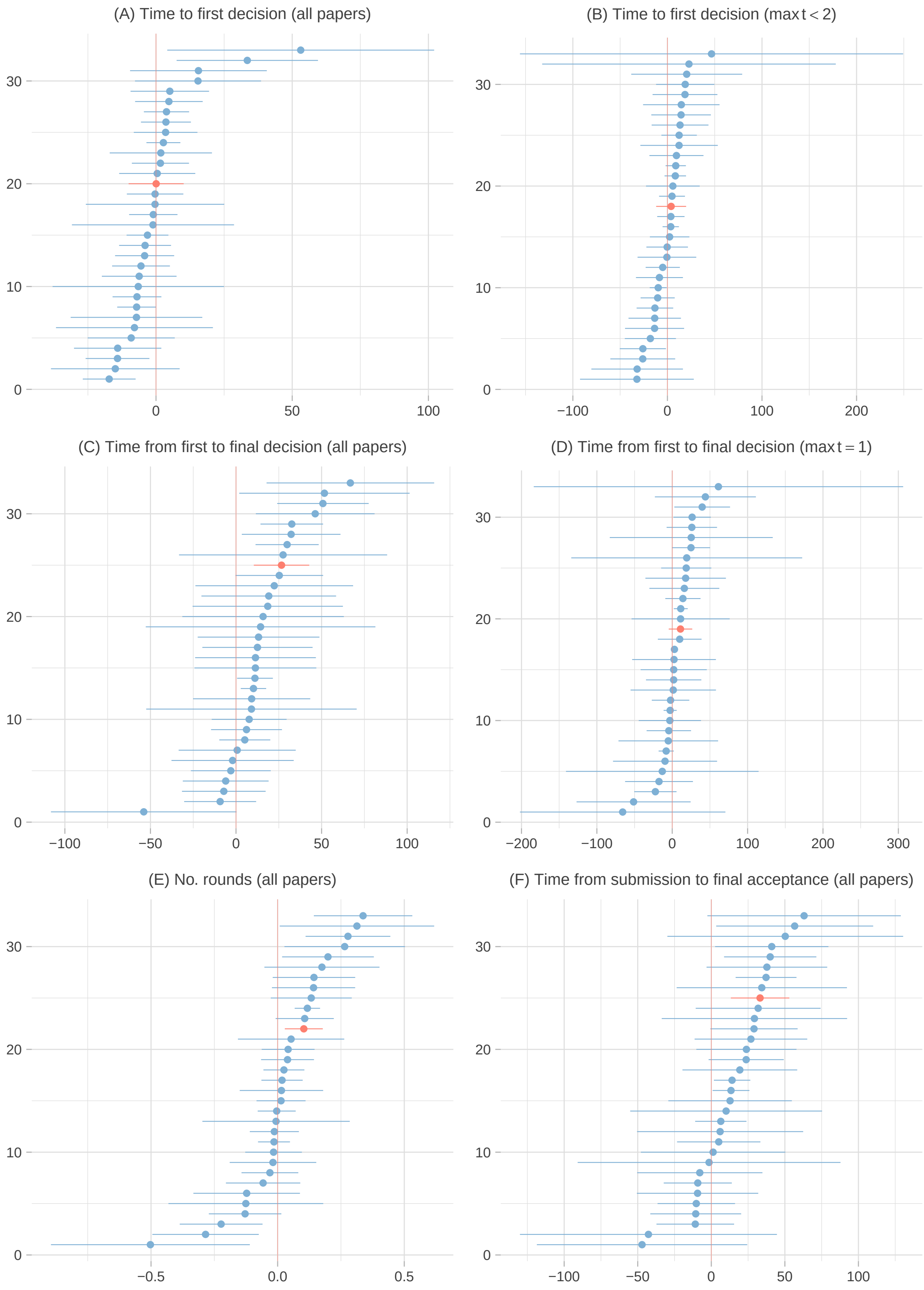

Figure B.2: Gender gap heterogeneity across journals

Note. Figures (A)–(E) plot the coefficient on female from regressions identical to those in the top panel Table 2 except that samples are restricted to observations from particular journals; Figure B.2F similarly plots the gender gap in total time under review from the same regression used to estimate column (1) of Table D.2 (Appendix D). Estimates for *Energy Economics* are highlighted in red. Bands represent 90% confidence intervals.

B.4 Author gender, total rounds and time to first decision

Table B.2: Correlation between rounds of review and time to first decision

	32 econ. and finance journals		*Energy Economics*	
	corr. author	solo author	corr. author	solo author
	(1)	(2)	(3)	(4)
female	7.007**	15.955***	18.783	61.979**
	(3.290)	(5.775)	(13.631)	(27.443)
max t	−7.852***	−7.188***	−14.570***	−7.272
	(1.916)	(2.320)	(3.099)	(7.604)
citations (asinh)	−4.899***	−5.334***	−27.417***	−27.740***
	(1.014)	(1.370)	(2.385)	(6.346)
female× max t	−3.438	−8.060**	−10.042	−40.207***
	(2.287)	(3.845)	(6.364)	(12.583)
No. obs.	24,511	6,532	2,359	447
R^2	0.170	0.170	0.115	0.135
Mean dep. var.	130.491	138.246	152.493	152.770
Journal	✓	✓		
Year	✓	✓	✓	✓

Note. Table displays the coefficients from regressing time to first decision on female authorship, total rounds of review (max t), citations (asinh), the interaction between max t and female authorship and year and journal fixed effects. Standard errors (in parentheses) clustered by journal and corresponding author country in columns (1) and (2) and by manuscript in columns (3) and (4). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table B.2 displays the coefficients from regressing time to first decision on female authorship, total rounds of review (max t), citations (asinh), the interaction between female and max t and year and journal fixed effects in the data for 32 economics and finance journals (first panel) and *Energy Economics* (second panel). In all four instances, the coefficient on female is negative or indistinguishable from zero when the interaction between female and max t is not taken into account.[5] But once it is, the coefficient on female is large and positive while the coefficient on the interaction is large and negative, especially among solo-authored papers (Table B.2).

This pattern may suggest that referees spend longer in the first round evaluating female-authored papers when they expect them to be (eventually) accepted; however, if they anticipate rejection or the outcome is less certain, they may spend less time on papers in the first round but more time on them in subsequent rounds. For example, suppose referees are more likely to initially reject female-authored papers, conditional on acceptance criteria. (Recall that referees spend less time reviewing papers that they reject compared to papers that they recommend revising (Appendix A.1).) When editors over-ride their rejections, these referees may, as a consequence, be more likely to ask for changes that require several rounds of revisions. See Card *et al.* (2020) for evidence that referees do indeed make less favourable recommendations for female-authored papers conditional on quality, but that editors' decisions partially correct for this bias.

[5]See column (1) in Table 2 and Figure B.2A in Appendix B.3 for results corresponding to columns (1)–(3) that do not control for max t or the interaction. The coefficient on female from a similar regression among solo-authored papers published in *Energy Economics* is –8.4 (standard error 12.7).

C Accounting for additional control variables

C.1 Editor fixed effects

Table C.1 and Figure C.1 replicate Table 1 and Figure 3, respectively, but include fixed effects for handling editors.

Results in Table C.1 and Figure C.1 are very similar to those in Table 1 and Figure 3, respectively.

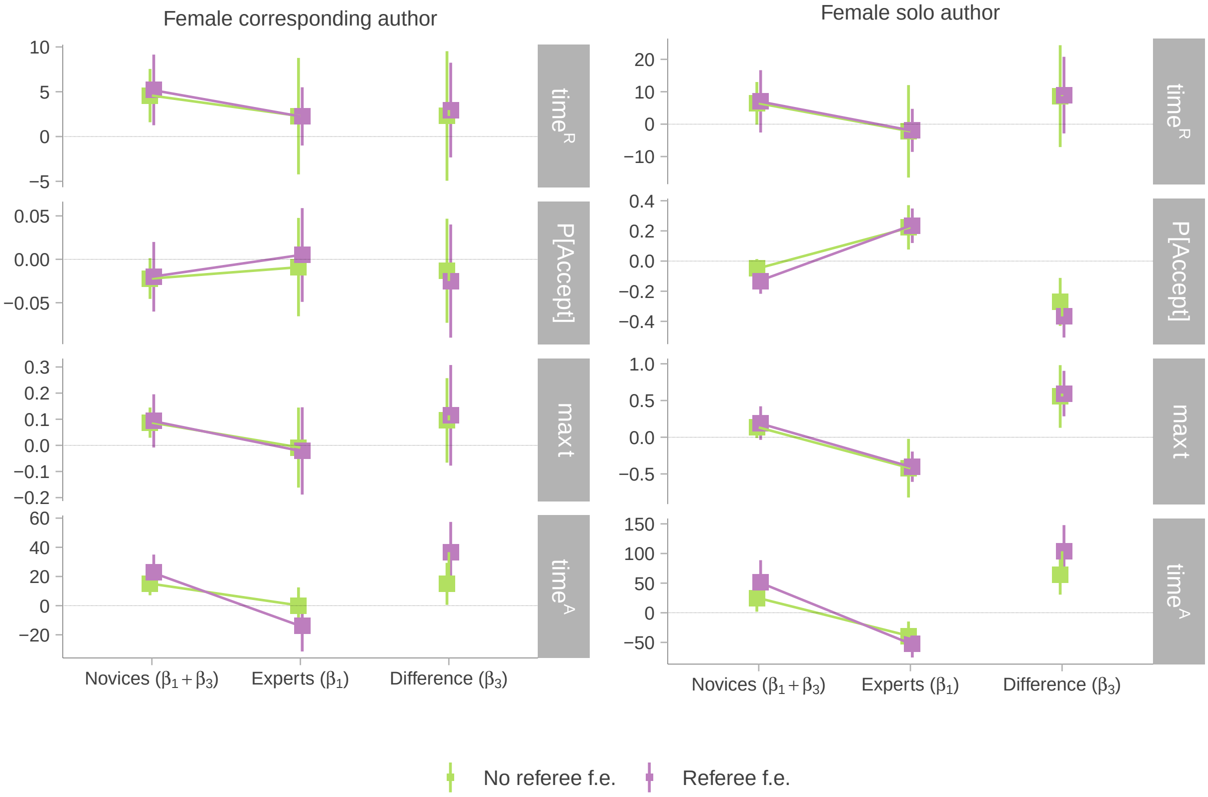

Figure C.1: Figure 3, controlling for editor fixed effects

Note. Graph is identical to Figure 3, except all results control for editor fixed effects.

Table C.1: Table 1, controlling for editor fixed effects

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	4.506***	4.642***	−0.026*	0.083*	10.623**	10.476*
	(1.592)	(1.608)	(0.014)	(0.044)	(5.259)	(5.418)
t (round)	−15.069***	−10.912***	0.364***		−42.603***	−28.316***
	(0.901)	(1.518)	(0.016)		(2.380)	(2.190)
citations (asinh)	−4.706***		0.004	−0.017	−12.261***	
	(0.837)		(0.004)	(0.014)	(1.796)	
R_{ijt} (referee's recommendation)						
revise (minor)		4.276**				
		(1.750)				
revise (major)		6.335***				
		(2.280)				
D_{it} (editor's decision)						
revise (major)		5.969*				3.059
		(3.106)				(7.468)
revise (minor)		4.119**				6.016
		(2.000)				(4.568)
$\mathcal{I}_j$ (prominence)					−1.161***	−1.459***
					(0.347)	(0.294)
D_{jt-1}						44.146***
						(4.218)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.090	0.082	0.277	0.097	0.116	0.137
Bounds (β_1)	[4.28, 4.51]	[4.56, 4.64]	[-0.03, -0.03]	[0.05, 0.08]	[8.57, 10.62]	[8.28, 10.48]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Solo-authored papers						
female	9.506**	8.907**	−0.008	0.060	28.807	28.222
	(4.026)	(4.115)	(0.036)	(0.125)	(17.492)	(17.789)
t (round)	−11.860***	−8.332***	0.327***		−34.792***	−27.267***
	(1.328)	(1.893)	(0.029)		(5.612)	(5.761)
citations (asinh)	−3.428***		0.000	−0.029	−4.689*	
	(1.273)		(0.009)	(0.040)	(2.728)	
R_{ijt} (referee's recommendation)						
revise (minor)		8.706**				
		(4.156)				
revise (major)		11.785**				
		(4.609)				
D_{it} (editor's decision)						
revise (major)		0.263				
		(6.303)				
revise (minor)		4.151				
		(5.172)				
$\mathcal{I}_j$ (prominence)					−3.996***	−3.710***
					(1.506)	(1.417)
D_{jt-1}						30.938***
						(8.746)
No. obs.	1,275	1,275	1,350	447	744	744
R^2	0.082	0.084	0.268	0.080	0.106	0.119
Bounds (β_1)	[9.51, 10.09]	[8.86, 8.91]	[-0.02, -0.01]	[0.05, 0.06]	[21.97, 28.81]	[20.78, 28.22]
Mean dep. var.	53.395	53.395	0.245	1.669	67.227	67.227
Year	✓	✓	✓	✓	✓	✓
Editor	✓	✓	✓	✓	✓	✓

Note. Estimates are identical to those in Table 1 except that all results control for editor fixed effects. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.2 Referee fixed effects

Table C.2 replicates Table 1 but includes fixed effects for each referee. As discussed in Section 4.3, the number of referees who reviewed solo-authored papers by both sexes is small, preventing us from reliably accounting for referee fixed effects when restricting the sample to solo-authored papers. The bottom panel of Table C.2 therefore compares solo female-authored papers to papers by male corresponding authors. Given the large number of fixed effects, we also do not estimate Oster bounds for β_1.

Although results in Table C.2 are somewhat more noisily estimated relative to those in Table 1, accounting for referee fixed effects does not appear to have a meaningful impact on the gender gap in time and length of peer review.

Table C.2: Table 1, controlling for referee fixed effects

	Eqn. (5) Days spent with referees		Eqn. (6) ℙ[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	3.795**	3.521**	−0.027	0.084	17.593	16.276
	(1.559)	(1.558)	(0.021)	(0.080)	(17.156)	(17.458)
t (round)	−13.847***	−10.068***	0.446***		−46.310***	−42.056***
	(1.361)	(2.242)	(0.017)		(7.044)	(7.802)
citations (asinh)	−3.313***		0.002	0.019	−11.782**	
	(1.103)		(0.005)	(0.026)	(4.600)	
R_{ijt} (referee's recommendation)						
revise (minor)		7.995***				
		(1.506)				
revise (major)		7.179***				
		(1.796)				
D_{it} (editor's decision)						
revise (major)		1.102				−2.344
		(3.222)				(14.869)
revise (minor)		0.836				−4.061
		(1.696)				(10.932)
$\mathcal{I}_j$ (prominence)					−0.977	−1.354*
					(0.739)	(0.760)
D_{jt-1}						29.445**
						(12.078)
No. obs.	7,044	7,044	7,473	4,724	2,866	2,866
R^2	0.628	0.628	0.552	0.548	0.510	0.508
Mean dep. var.	53.280	53.280	0.272	1.684	97.371	97.371
Solo-authored papers						
female	5.860	5.211	−0.076*	0.083	53.549	52.181
	(4.574)	(4.638)	(0.044)	(0.172)	(72.394)	(72.658)
t (round)	−14.006***	−9.788***	0.448***		−49.971***	−44.167***
	(1.422)	(2.304)	(0.019)		(9.934)	(9.675)
citations (asinh)	−2.873**		−0.002	0.000	−8.144*	
	(1.229)		(0.006)	(0.033)	(4.879)	
R_{ijt} (referee's recommendation)						
revise (minor)		8.371***				
		(1.805)				
revise (major)		8.071***				
		(2.248)				
D_{it} (editor's decision)						
revise (major)		1.371				
		(3.561)				
revise (minor)		1.159				
		(2.010)				
$\mathcal{I}_j$ (prominence)					−0.405	−0.560
					(1.039)	(0.999)
D_{jt-1}						37.053***
						(12.796)
No. obs.	4,890	4,890	5,188	3,293	1,985	1,985
R^2	0.634	0.635	0.576	0.608	0.539	0.541
Mean dep. var.	53.386	53.386	0.273	1.679	97.656	97.656
Year	✓	✓	✓	✓	✓	✓
Referee	✓	✓	✓	✓	✓	✓

Note. Estimates are identical to those in Table 1 except that (a) all results control for referee fixed effects; and (b) solo female-authored papers are compared to papers by male corresponding authors (see Section 4.3 for justification). ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.3 *JEL* code fixed effects

Table C.3 and Figure C.2 replicate Table 1 and Figure 3 but include fixed effects for secondary *JEL* codes. Because few codes contain a non-negligible number of solo-authored papers by both men and women, the bottom panel of Table C.3 compares solo-female authored papers to papers by male corresponding authors. (Results for solo-female-authored papers in Figure 3 are already relative to papers by male corresponding authors, see Section 4.3.)

Results in Table C.3 and Figure C.2 are similar to those in Table 1 and Figure 3, although the coefficient on female is smaller and more noisily estimated in the bottom panel of the former.

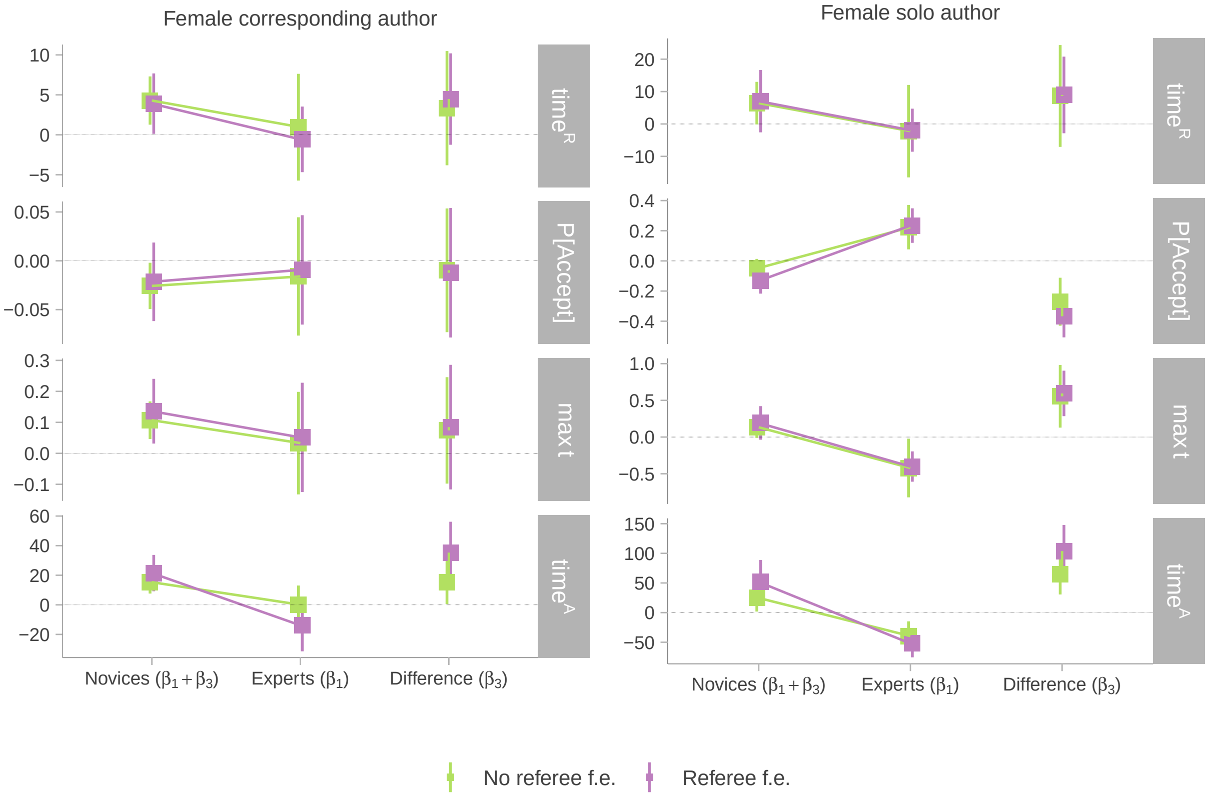

Figure C.2: Figure 3, controlling for secondary *JEL* code fixed effects

Note. Graph is identical to Figure 3, except all results control for editor fixed effects.

Table C.3: Table 1, controlling for secondary *JEL* code fixed effects

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	3.929**	4.119**	−0.030**	0.101**	11.015**	10.602**
	(1.674)	(1.675)	(0.014)	(0.048)	(5.142)	(5.192)
t (round)	−16.082***	−13.222***	0.369***		−42.793***	−28.924***
	(0.977)	(1.560)	(0.016)		(2.398)	(2.199)
citations (asinh)	−4.716***		0.002	0.011	−11.253***	
	(0.710)		(0.004)	(0.016)	(1.838)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.762***				
		(1.836)				
revise (major)		7.217***				
		(2.282)				
D_{it} (editor's decision)						
revise (major)		1.346				0.655
		(3.030)				(7.297)
revise (minor)		1.754				4.971
		(1.953)				(4.552)
$\mathcal{I}_j$ (prominence)					−1.023***	−1.135***
					(0.318)	(0.300)
D_{jt-1}						44.261***
						(4.223)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.116	0.108	0.291	0.094	0.152	0.176
Bounds (β_1)	[3.03, 3.93]	[3.43, 4.12]	[-0.04, -0.03]	[0.09, 0.10]	[9.19, 11.02]	[8.30, 10.60]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Solo-authored papers						
female	6.775*	6.281	−0.043	0.109	18.392	16.486
	(3.948)	(4.086)	(0.037)	(0.124)	(15.302)	(15.442)
t (round)	−16.017***	−12.229***	0.365***		−41.673***	−28.243***
	(1.052)	(1.574)	(0.017)		(2.892)	(2.708)
citations (asinh)	−4.753***		0.007	−0.006	−11.702***	
	(0.819)		(0.005)	(0.020)	(2.170)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.410**				
		(2.103)				
revise (major)		8.153***				
		(2.566)				
D_{it} (editor's decision)						
revise (major)		3.630				
		(3.311)				
revise (minor)		3.628				
		(2.357)				
$\mathcal{I}_j$ (prominence)					−1.096***	−1.293***
					(0.387)	(0.359)
D_{jt-1}						45.088***
						(4.301)
No. obs.	4,890	4,890	5,188	1,651	2,666	2,666
R^2	0.123	0.116	0.293	0.113	0.164	0.184
Bounds (β_1)	[5.24, 6.77]	[4.13, 6.28]	[-0.06, -0.04]	[0.10, 0.11]	[6.95, 18.39]	[2.78, 16.49]
Mean dep. var.	53.386	53.386	0.273	1.618	70.806	70.806
Year	✓	✓	✓	✓	✓	✓
JEL (secondary)	✓	✓	✓	✓	✓	✓

Note. Estimates are identical to those in Table 1 except that (a) all results control for secondary *JEL* code fixed effects; and (b) solo female-authored papers are compared to papers by male corresponding authors. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

C.4 Authors' institutional rank

Table C.4 and Figure C.3 replicate Table 1 and Figure 3 but include fixed effects for the rank of authors' institutions in the fields of energy economics and finance. 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.

Results in Table C.4 and Figure C.3 are similar to those reported in Table 1 and Figure 3.

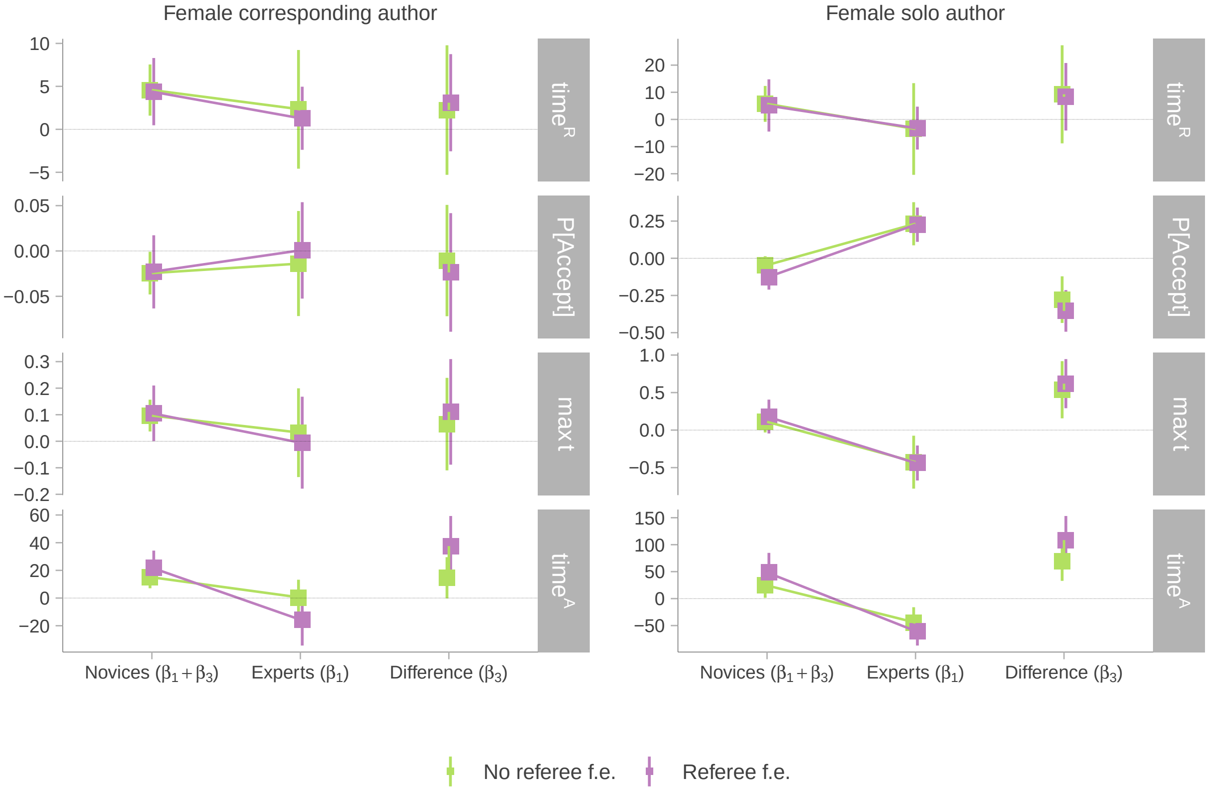

Figure C.3: Figure 3, controlling for institutional rank fixed effects

Note. Graph is identical to Figure 3, except all results control for institutional rank fixed effects.

Table C.4: Table 1, controlling for authors' institutional rank

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	4.392***	4.467***	−0.028**	0.096**	10.645**	10.311*
	(1.629)	(1.648)	(0.014)	(0.046)	(5.338)	(5.515)
t (round)	−15.896***	−12.723***	0.363***		−41.906***	−28.839***
	(0.969)	(1.611)	(0.016)		(2.341)	(2.177)
citations (asinh)	−5.140***		0.002	0.017	−11.388***	
	(0.794)		(0.004)	(0.015)	(1.830)	
R_{ijt} (referee's recommendation)						
revise (minor)		5.462***				
		(1.792)				
revise (major)		7.173***				
		(2.276)				
D_{it} (editor's decision)						
revise (major)		2.738				2.487
		(3.175)				(7.482)
revise (minor)		2.649				5.403
		(1.978)				(4.505)
$\mathcal{I}_j$ (prominence)					−1.014***	−1.156***
					(0.345)	(0.292)
D_{jt-1}						43.528***
						(4.204)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.083	0.072	0.275	0.042	0.115	0.138
Bounds (β_1)	[4.06, 4.39]	[4.21, 4.47]	[-0.04, -0.03]	[0.08, 0.10]	[8.65, 10.64]	[7.98, 10.31]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Solo-authored papers						
female	9.280**	8.550**	−0.005	0.043	29.518*	28.848*
	(4.150)	(4.256)	(0.035)	(0.122)	(17.058)	(17.354)
t (round)	−12.228***	−8.640***	0.331***		−35.272***	−27.822***
	(1.298)	(1.810)	(0.029)		(5.473)	(5.640)
citations (asinh)	−4.097***		0.001	−0.015	−5.257**	
	(1.288)		(0.009)	(0.041)	(2.654)	
R_{ijt} (referee's recommendation)						
revise (minor)		9.787**				
		(4.226)				
revise (major)		12.682***				
		(4.557)				
D_{it} (editor's decision)						
revise (major)		−0.749				
		(6.116)				
revise (minor)		3.906				
		(5.130)				
$\mathcal{I}_j$ (prominence)					−4.043**	−3.864**
					(1.591)	(1.530)
D_{jt-1}						30.214***
						(8.657)
No. obs.	1,275	1,275	1,350	447	744	744
R^2	0.070	0.068	0.263	0.066	0.104	0.115
Bounds (β_1)	[9.28, 9.63]	[8.14, 8.55]	[-0.01, -0.01]	[0.01, 0.04]	[23.45, 29.52]	[22.08, 28.85]
Mean dep. var.	53.395	53.395	0.245	1.669	67.227	67.227
Year	✓	✓	✓	✓	✓	✓
Inst. rank	✓	✓	✓	✓	✓	✓

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

C.5 Author prominance

Table C.5 and Figure C.4 replicate columns (1)–(4) in Table 1 and the first three rows in Figure 3 but additionally control for the "prominence" of the most prominent co-author on a paper.[6] To proxy for "prominence", we total the number of papers an author previously published in *Energy Economics* at the time his manuscript is submitted.

Table C.6 similarly replicates columns (1)–(3) in Table 2 controlling for the prominence of the most prominent co-author, where "prominence" is now defined as the total number of papers an author previously published in any of the 32 economics and finance journals included in the estimation sample.[7]

The gender gaps in reviewing and revising times are largely robust to controlling for author prominence, although results for solo-authored papers in Table C.5 are more noisily estimated, especially in columns (3) and (4). The coefficient on author prominence in both Tables C.5 and C.6 suggest that referees are quicker to review and accept papers by more prominent authors and more prominent authors are faster to revise their papers.

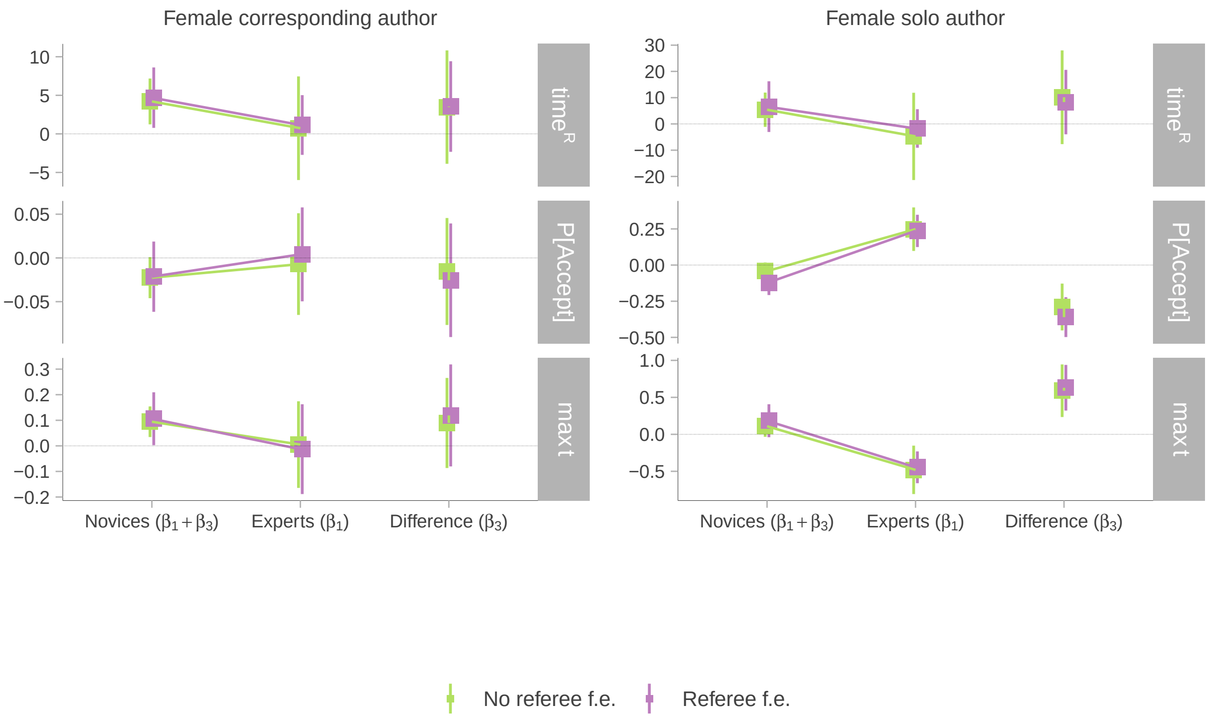

Figure C.4: Figure 3, controlling for "prominence" of the most prominent co-author

Note. Graphs are identical to the first three rows of graphs in Figure 3, except that all results control for the prominence of the most prominent co-author on a paper.

[6]Columns (5)–(6) of Table 1 and the final row of graphs in Figure 3 already control for co-author prominence and are therefore omitted from Table C.5 and Figure C.4.

[7]Results are similar if we instead proxy for author prominence with the number of times an author had previously published in the specific journal he submitted his manuscript to (not shown).

Table C.5: Table 1, controlling for the "prominence" of the most prominent co-author

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds
	(1)	(2)	(3)	(4)
Female corresponding authors				
female	3.752**	3.720**	−0.026*	0.096**
	(1.617)	(1.635)	(0.014)	(0.046)
t (round)	−15.964***	−13.143***	0.363***	
	(0.967)	(1.574)	(0.016)	
citations (asinh)	−4.579***		0.000	0.018
	(0.780)		(0.004)	(0.015)
$\mathcal{I}_j$ (prominence)	−0.697***	−0.955***	0.003**	−0.009***
	(0.187)	(0.192)	(0.001)	(0.003)
R_{ijt} (referee's recommendation)				
revise (minor)		5.669***		
		(1.774)		
revise (major)		7.168***		
		(2.267)		
D_{it} (editor's decision)				
revise (major)		1.655		
		(3.085)		
revise (minor)		2.290		
		(1.949)		
No. obs.	7,044	7,044	7,473	2,359
R^2	0.087	0.079	0.276	0.040
Bounds (β_1)	[2.76, 3.75]	[2.70, 3.72]	[-0.03, -0.03]	[0.08, 0.10]
Mean dep. var.	53.280	53.280	0.272	1.622
Solo-authored papers				
female	8.497**	7.610*	−0.002	0.034
	(4.013)	(4.115)	(0.035)	(0.125)
t (round)	−12.383***	−9.450***	0.331***	
	(1.306)	(1.808)	(0.029)	
citations (asinh)	−3.600***		−0.003	−0.006
	(1.237)		(0.009)	(0.041)
$\mathcal{I}_j$ (prominence)	−1.181**	−1.543***	0.013***	−0.045***
	(0.476)	(0.488)	(0.004)	(0.012)
R_{ijt} (referee's recommendation)				
revise (minor)		9.629**		
		(4.187)		
revise (major)		12.870***		
		(4.515)		
D_{it} (editor's decision)				
revise (major)		−2.453		
		(6.210)		
revise (minor)		3.561		
		(5.139)		
No. obs.	1,275	1,275	1,350	447
R^2	0.073	0.075	0.264	0.057
Bounds (β_1)	[8.03, 8.50]	[6.24, 7.61]	[-0.01, 0.00]	[-0.01, 0.03]
Mean dep. var.	53.395	53.395	0.245	1.669
Year	✓	✓	✓	✓

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

Table C.6: Table 2, controlling for the "prominence" of the most prominent co-author

	Time to first decision		No. rounds (max t)
	All papers	max $t \leq 1$	All papers
	(1)	(2)	(3)
Female corresponding authors			
female	0.563	3.508**	0.038*
	(0.350)	(1.541)	(0.021)
citations (asinh)	−5.085***	−4.460***	0.018**
	(0.983)	(1.384)	(0.008)
prominence	0.201	−0.076	−0.008***
	(0.210)	(0.241)	(0.001)
No. obs.	24,511	10,708	24,560
R^2	0.166	0.210	0.229
Bounds (female)	[0.56, 3.74]	[3.51, 6.31]	[0.04, 0.04]
Mean dep. var.	130.491	127.557	1.779
Solo-authored papers			
female	0.570	8.851*	0.077***
	(2.915)	(4.663)	(0.024)
citations (asinh)	−5.598***	−5.594***	0.030***
	(1.370)	(1.928)	(0.007)
prominence	0.751	0.781	−0.004
	(0.908)	(0.873)	(0.003)
No. obs.	6,532	2,876	6,553
R^2	0.166	0.233	0.261
Bounds (female)	[0.57, 3.00]	[8.26, 8.85]	[0.07, 0.08]
Mean dep. var.	138.246	135.225	1.778
Journal	✓	✓	✓
Year	✓	✓	✓

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

C.6 Number of co-authors

Tables C.7 and C.8 and Figure C.5 replicate the top halves of Tables 1 and 2 and Figure 3, respectively, but control for the number of co-authors on a paper. Results are very similar those reported in Tables 1 and 2 and Figure 3.

Table C.7: Table 1 (top panel), controlling for number of co-authors

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
female	4.536***	4.794***	−0.028**	0.103**	11.192**	10.950**
	(1.648)	(1.678)	(0.014)	(0.046)	(5.235)	(5.391)
t (round)	−15.886***	−12.604***	0.363***		−41.691***	−28.574***
	(0.972)	(1.624)	(0.016)		(2.355)	(2.196)
citations (asinh)	−5.316***		0.002	0.010	−11.829***	
	(0.843)		(0.004)	(0.015)	(1.798)	
no. co-authors	0.963*	0.607	0.004	0.002	3.326**	2.656*
	(0.506)	(0.508)	(0.004)	(0.013)	(1.579)	(1.599)
R_{ijt} (referee's recommendation)						
revise (minor)		5.441***				
		(1.795)				
revise (major)		7.258***				
		(2.277)				
D_{it} (editor's decision)						
revise (major)		2.891				2.402
		(3.216)				(7.480)
revise (minor)		2.598				5.490
		(1.989)				(4.504)
$\mathcal{I}_j$ (prominence)					−1.385***	−1.585***
					(0.339)	(0.267)
D_{jt-1}						43.790***
						(4.211)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.084	0.071	0.275	0.038	0.115	0.137
Bounds (β_1)	[4.35, 4.54]	[4.79, 4.86]	[-0.04, -0.03]	[0.10, 0.10]	[9.76, 11.19]	[9.27, 10.95]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Year	✓	✓	✓	✓	✓	✓

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

Table C.8: Table 2 (top panel), controlling for number of co-authors

	Time to first decision		No. rounds (max t)	Time from first to final decision	
	All papers	max $t \leq 1$	All papers	All papers	max $t = 1$
	(1)	(2)	(3)	(4)	(5)
female	0.419	3.490**	0.043**	14.483***	5.641**
	(0.453)	(1.699)	(0.020)	(3.016)	(2.390)
citations (asinh)	−5.124***	−4.551***	0.017*	−12.295***	−4.468***
	(0.947)	(1.339)	(0.008)	(2.840)	(1.626)
no. co-authors	0.647	0.343	−0.014	8.463***	4.555**
	(1.011)	(1.440)	(0.009)	(2.169)	(1.943)
max t				93.657***	
				(8.073)	
prominence				−1.263***	−1.076***
				(0.400)	(0.319)
No. obs.	24,499	10,701	24,548	23,896	10,077
R^2	0.166	0.210	0.228	0.409	0.248
Bounds (female)	[0.42, 3.49]	[3.49, 6.28]	[0.04, 0.04]	[11.05, 14.48]	[4.93, 5.64]
Mean dep. var.	130.496	127.560	1.779	214.964	112.291
Journal	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓

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

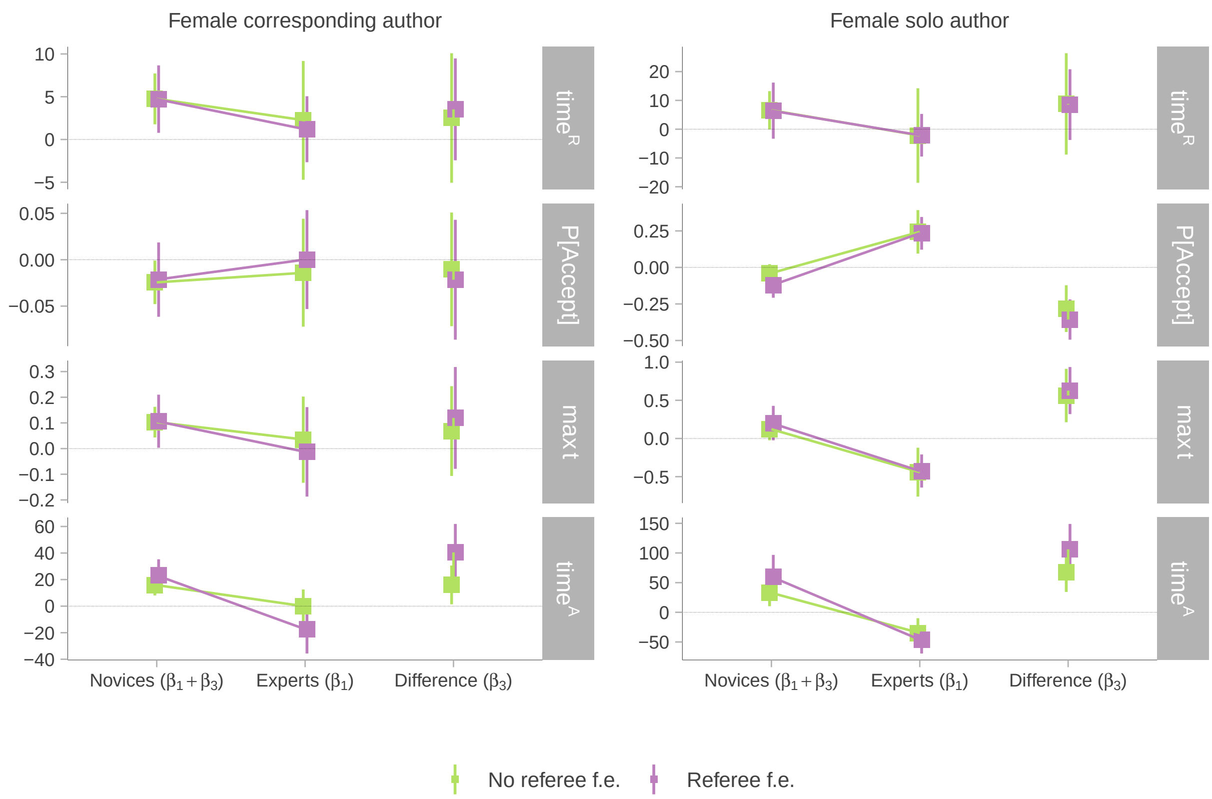

Figure C.5: Figure 3, controlling for number of co-authors

Note. Graph is identical to Figure 3, except all results control for number of co-authors.

C.7 Manuscript length

Tables C.9, C.10 and C.11 and Figure C.6 replicate the top half of Table 1, the bottom half of Table 1, Table 2 and Figure 3, respectively, but control for manuscript length. Results are very similar to those reported in Tables 1 and 2 and Figure 3.

Table C.9: Table 1 (top panel), controlling for manuscript length

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
female	4.223**	4.586***	−0.027*	0.099**	10.767**	10.622**
	(1.670)	(1.703)	(0.014)	(0.046)	(5.189)	(5.362)
t (round)	−15.930***	−12.710***	0.363***		−41.960***	−28.801***
	(0.970)	(1.633)	(0.016)		(2.376)	(2.214)
citations (asinh)	−5.369***		0.003	0.008	−11.830***	
	(0.825)		(0.004)	(0.015)	(1.804)	
no. pages	0.521***	0.379*	−0.003**	0.013***	1.031**	0.818**
	(0.195)	(0.208)	(0.001)	(0.005)	(0.427)	(0.405)
R_{ijt} (referee's recommendation)						
revise (minor)		5.376***				
		(1.797)				
revise (major)		7.128***				
		(2.278)				
D_{it} (editor's decision)						
revise (major)		2.755				1.842
		(3.252)				(7.464)
revise (minor)		2.460				5.170
		(1.998)				(4.500)
$\mathcal{I}_j$ (prominence)					−1.284***	−1.504***
					(0.396)	(0.292)
D_{jt-1}						43.960***
						(4.206)
No. obs.	7,044	7,044	7,473	2,359	3,819	3,819
R^2	0.085	0.072	0.276	0.042	0.115	0.137
Bounds (β_1)	[3.72, 4.22]	[4.45, 4.59]	[-0.03, -0.03]	[0.09, 0.10]	[8.90, 10.77]	[8.61, 10.62]
Mean dep. var.	53.280	53.280	0.272	1.622	71.209	71.209
Year	✓	✓	✓	✓	✓	✓

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

Table C.10: Table 1 (bottom panel), controlling for manuscript length

	Eqn. (5) Days spent with referees		Eqn. (6) $\mathbb{P}$[Accept]	Eqn. (7) No. rounds	Eqn. (8) Days spent with authors	
	(1)	(2)	(3)	(4)	(5)	(6)
female	9.608**	8.847**	−0.013	0.074	30.348*	29.721*
	(3.980)	(4.107)	(0.035)	(0.123)	(17.106)	(17.372)
t (round)	−12.244***	−8.661***	0.329***		−34.949***	−27.589***
	(1.311)	(1.853)	(0.029)		(5.442)	(5.624)
citations (asinh)	−4.172***		0.003	−0.024	−4.979*	
	(1.261)		(0.009)	(0.040)	(2.713)	
no. pages	0.149	−0.025	−0.002	0.010	−0.443	−0.344
	(0.260)	(0.266)	(0.002)	(0.009)	(0.720)	(0.685)
R_{ijt} (referee's recommendation)						
revise (minor)		9.763**				
		(4.229)				
revise (major)		12.817***				
		(4.585)				
D_{it} (editor's decision)						
revise (major)		−0.809				
		(6.195)				
revise (minor)		3.908				
		(5.130)				
$\mathcal{I}_j$ (prominence)					−3.949**	−3.703**
					(1.600)	(1.503)
D_{jt-1}						30.152***
						(8.680)
No. obs.	1,275	1,275	1,350	447	744	744
R^2	0.069	0.067	0.260	0.050	0.102	0.114
Bounds (β_1)	[9.61, 10.27]	[8.75, 8.85]	[-0.03, -0.01]	[0.07, 0.07]	[25.19, 30.35]	[23.92, 29.72]
Mean dep. var.	53.395	53.395	0.245	1.669	67.227	67.227
Year	✓	✓	✓	✓	✓	✓

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

Table C.11: Table 2, controlling for manuscript length

	Time to first decision		No. rounds (max t)	Time from first to final decision	
	All papers	max $t \leq 1$	All papers	All papers	max $t = 1$
	(1)	(2)	(3)	(4)	(5)
Female corresponding authors					
female	0.325	3.487**	0.042**	13.638***	4.873**
	(0.369)	(1.681)	(0.021)	(2.754)	(2.217)
citations (asinh)	−5.070***	−4.512***	0.014*	−11.771***	−4.243**
	(1.008)	(1.432)	(0.008)	(2.827)	(1.629)
no. pages	0.269	0.169	0.003	2.522***	2.364***
	(0.310)	(0.297)	(0.003)	(0.662)	(0.316)
max t				93.369***	
				(8.163)	
prominence				−0.721**	−0.784***
				(0.347)	(0.271)
No. obs.	24,511	10,708	24,560	23,907	10,083
R^2	0.166	0.210	0.229	0.410	0.253
Bounds (female)	[0.32, 3.34]	[3.49, 6.27]	[0.04, 0.04]	[9.85, 13.64]	[3.94, 4.87]
Mean dep. var.	130.491	127.557	1.779	214.954	112.311
Solo-authored papers					
female	−0.109	8.324	0.078***	20.446***	11.295
	(3.077)	(5.086)	(0.023)	(5.142)	(7.198)
citations (asinh)	−5.649***	−5.538***	0.029***	−12.335***	−4.823**
	(1.374)	(1.902)	(0.007)	(3.133)	(2.128)
no. pages	0.622*	0.266	0.001	2.747***	2.907***
	(0.321)	(0.334)	(0.004)	(0.882)	(0.590)
max t				95.918***	
				(10.293)	
prominence				−6.012***	−3.474***
				(1.528)	(0.815)
No. obs.	6,532	2,876	6,553	6,380	2,715
R^2	0.167	0.233	0.261	0.424	0.281
Bounds (female)	[-0.11, 1.83]	[7.31, 8.32]	[0.07, 0.08]	[8.66, 20.45]	[8.81, 11.29]
Mean dep. var.	138.246	135.225	1.778	208.014	102.014
Journal	✓	✓	✓	✓	✓
Year	✓	✓	✓	✓	✓

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

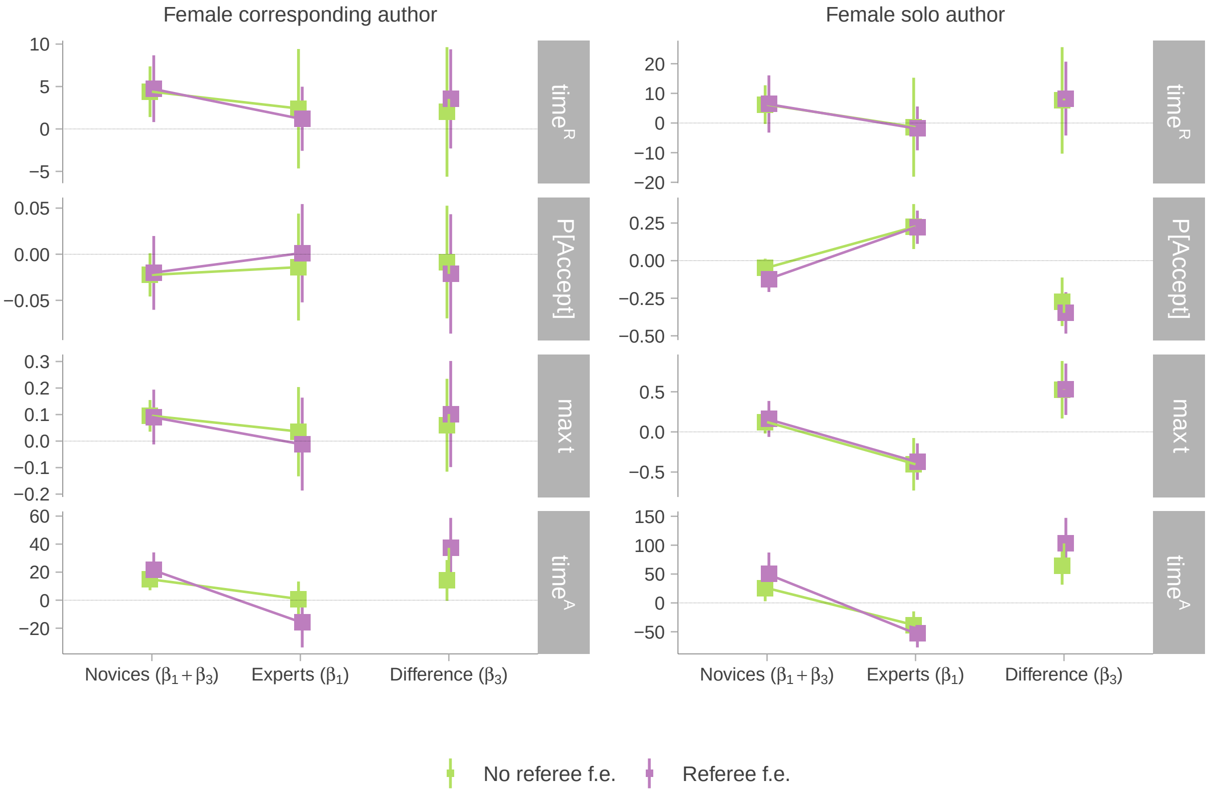

Figure C.6: Figure 3, controlling for manuscript length

Note. Graph is identical to Figure 3, except all results control for manuscript length.

D Cumulative gender gaps in time spent in review

Tables D.1 and D.2 estimate the gender gap in the total number of days spent under review at *Energy Economics* and the additional 32 economics and finance journals, respectively.

On average, papers published in *Energy Economics* go through about two rounds of review (see Section 2), so estimates in Table D.1 are unsurprisingly about twice the size of corresponding round-specific estimates in Table 1. The first two horizontal panels in Table D.1 show the aggregate number of days manuscripts spend with each referee and with authors.[8] They suggest that referees take, in total, 7–14 days longer evaluating female-authored papers, and female authors spend around 21–59 days longer revising their papers. The final two columns of Table D.1 show gender gaps in the difference between submission and final acceptance dates; according to these estimates, female authors spend around 29–60 days longer under review.

Table D.2 displays results from regressing the difference between submission and final acceptance dates on citations (asinh), author prominence, journal and year fixed effects, corresponding authors' genders (column (1)) and solo authors' genders (column (2)). These estimates suggest that female-authored papers spend, in total, 18–29 days longer under review.

[8] "Days with referee" aggregates the time each referee spends evaluating an eventually accepted manuscript irrespective of his own recommendation. It therefore also includes time spent with referees who recommended rejection in a given round, in contrast to the sample used in Table 1.

Table D.1: Gender gaps in cumulative time under review at *Energy Economics*

	Days with referee		Days with authors		Total days	
	(1)	(2)	(3)	(4)	(5)	(6)
Female corresponding authors						
female	8.826***	7.429***	24.042***	21.120**	33.162**	29.375**
	(2.807)	(2.686)	(9.022)	(9.175)	(13.119)	(13.822)
citations (asinh)	−7.300***		−19.124***		−53.058***	
	(1.326)		(3.083)		(4.666)	
R_{ij0} (referee's initial recommendation)						
revise (minor)		7.787*				
		(4.183)				
revise (major)		42.287***				
		(4.555)				
reject		21.957***				
		(6.458)				
D_{i0} (editor's initial decision)						
revise (major)		−66.370*		135.266***		131.223***
		(39.672)		(7.865)		(30.649)
revise (minor)		−73.632*		70.760***		15.228
		(38.500)		(7.892)		(32.084)
$\mathcal{I}_j$ (prominence)			−1.853***	−2.187***	−4.306***	−6.023***
			(0.494)	(0.394)	(1.059)	(0.762)
No. obs.	4,724	4,724	2,359	2,359	2,359	2,359
R^2	0.029	0.099	0.062	0.085	0.143	0.126
Bounds (β_1)	[7.8, 8.8]	[5.0, 7.4]	[19.1, 24.0]	[13.2, 21.1]	[19.3, 33.2]	[11.8, 29.4]
Mean dep. var.	84.894	84.894	115.382	115.382	337.843	337.843
Solo-authored papers						
female	12.739*	13.576*	59.235**	58.235*	59.587	49.824
	(6.960)	(7.029)	(29.295)	(29.754)	(36.243)	(39.067)
citations (asinh)	−5.459***		−10.063**		−45.859***	
	(2.056)		(4.910)		(9.227)	
R_{ij0} (referee's initial recommendation)						
revise (minor)		25.579***				
		(7.598)				
revise (major)		62.752***				
		(8.770)				
reject		51.362***				
D_{i0} (editor's initial decision)						
		(12.892)				
revise (major)		−35.810		104.856***		213.929***
		(49.163)		(19.575)		(78.801)
revise (minor)		−35.970		62.955***		135.189*
		(49.552)		(23.035)		(80.228)
$\mathcal{I}_j$ (prominence)			−5.607**	−5.182**	−11.883***	−14.308***
			(2.241)	(2.102)	(3.484)	(2.927)
No. obs.	869	869	447	447	447	447
R^2	0.028	0.123	0.092	0.104	0.158	0.128
Bounds (β_1)	[12.7, 13.1]	[13.6, 14.8]	[53.9, 59.2]	[51.8, 58.2]	[54.4, 59.6]	[34.6, 49.8]
Mean dep. var.	83.877	83.877	111.895	111.895	339.114	339.114
Year	✓	✓	✓	✓	✓	✓

Note. Figures correspond to coefficients from regressing total days spent with each referee (columns(1)–(2)), total days spent with authors (columns (3)–(4)) and the difference between a paper's submission and final acceptance dates (columns (5)–(6)) on the listed variables. In the top panel, we proxy for gender using the gender of the corresponding author; in the bottom panel, we compare solo-female-authored papers to solo-male-authored papers. Standard errors clustered by referee (columns (1)–(2)) and authors (columns (3)–(4)) in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

Table D.2: Gender gaps in cumulative time under review at 32 economics and finance journals

	Female corr. author	Female solo-authored
	(1)	(2)
female	18.191***	29.416***
	(3.519)	(5.922)
citations (asinh)	−14.360***	−14.133***
	(3.622)	(3.775)
prominence	−1.322**	−5.750***
	(0.494)	(1.783)
No. obs.	24,510	6,532
R^2	0.306	0.310
Bounds (female)	[18.11, 18.19]	[23.78, 29.42]
Mean dep. var.	339.815	340.970
Journal	✓	✓
Year	✓	✓

Note. Figures correspond to coefficients from regressing the difference between a paper's submission and final acceptance dates on the listed variables. In column (1), we proxy for gender using the gender of the corresponding author; in column (2), we compare solo-female-authored papers to solo-male-authored papers. Standard errors clustered by journal and authors' countries in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

E Gender gaps in citations

Table E.1 displays coefficients from a regression of citations (asinh) on female authorship, co-author prominence, number of authors and journal and year fixed effects. (See Hengel and Moon (2022) for an explanation and justification of this specification.) Results analysing data for *Energy Economics* and the additional 32 economics and finance journals are shown in the first and second panels, respectively. In columns (1) and (3), female authorship is defined as having a female corresponding author; in columns (2) and (4) we compare solo female-authored papers to solo male-authored papers.

Table E.1: Gender gaps in citations (asinh)

	Energy Economics		32 econ. and finance journals	
	fem. corr.	fem. solo	fem. corr	fem. solo
	(1)	(2)	(3)	(4)
female	−0.037	0.287*	0.040	0.057
	(0.062)	(0.156)	(0.027)	(0.033)
$\mathcal{I}_j$ (prominence)	0.060***	0.120***	0.028***	0.043***
	(0.006)	(0.024)	(0.003)	(0.010)
no. authors	0.041**		0.209***	
	(0.020)		(0.021)	
No. obs.	2,359	447	24,548	6,553
R^2	0.486	0.423	0.278	0.245
Mean dep. var.	2.653	2.670	2.846	2.444
Journal			✓	✓
Year	✓	✓	✓	✓

Note. Table displays coefficients from a regression of citations (asinh) on female authorship, co-author prominence, number of authors and journal and year fixed effects. Standard errors clustered by manuscript (first panel) and journal and author country (second panel) in parentheses. ***, ** and * statistically significant at 1%, 5% and 10%, respectively.

With the exception of column (1), results in Table E.1 suggest that female-authored papers are cited more than male-authored papers. At *Energy Economics*, the gap is 29 log points among solo-authored papers. At the other journals, it is 4–6 log points.

References

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

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.

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

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

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

Hengel, E. and E. Moon (2022). "Gender and quality at top economics journals". Mimeo.

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