Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

CHARITY AND FAVORITISM IN THE FIELD:

ARE FEMALE ECONOMISTS NICER (TO EACH OTHER)?

Jason Abrevaya and Daniel S. Hamermesh*

*Professor of Economics, University of Texas at Austin; Sue Killam Professor in the Foundations of Economics, University of Texas at Austin, and professor of labor economics, Maastricht University, IZA and NBER. We thank Brian Keeling for having collected and provided the data, Sandra Black, Joan Muysken, Stephen Trejo, participants at seminars at several universities and a referee for their suggestions.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

ABSTRACT

Using a very large sample of matched author-referee pairs, we examine how referees' and authors' genders affect the former's recommendations. Relying on changing author-referee matches, we find no evidence of gender differences among referees in charitableness; nor is there any effect of the interaction between the referees' and authors' genders. With substantial laboratory research showing gender differences in fairness, the results suggest that outside the laboratory an ethos of objectivity can overcome possible tendencies toward same-group favoritism/opposite-group discrimination.

ECON LIT CODES: J71

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

I. Introduction

Discrimination is perhaps the most heavily researched topic in the field of labor economics and perhaps even among all endeavors in applied economics. Much less has been done on differences in fairness/charitableness by individuals with different characteristics, but that too is attracting increasing attention (e.g., Andreoni and Vesterlund, 2001, and the Croson and Gneezy, 2009, survey). Very little research has combined these two topics, asking whether the amount of favoritism/discrimination varies with the extent of the match between the parties (but see Parsons et al, 2011, Price and Wolfers, 2010, for differences by race/ethnicity, and Dillingham et al, 1994, for some sparse evidence on gender).

Our purpose here is to combine these two questions, focusing on differences by gender. We ask whether women are more or less generous than men in making up-or-down recommendations about others' work, and whether their degree of generosity is affected by the gender of those whose output they are asked to judge. While we despair of distinguishing favoritism toward one's own group from discrimination against another, our results do provide some evidence on whether the extent of favoritism/discrimination differs by gender of the favoring/discriminating party.

Perhaps analogously to the theory of religious sects (Iannaccone, 1992), one might argue that the degree of solidarity within a group is a function of its relative size—that smaller groups will be more cohesive and more likely to favor other members of the group. We test this possibility too, examining whether women favor other women more when they account for a smaller share of the relevant population of those making the judgments or being judged.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

II. Modeling Preferences and Favoritism

We seek to model interactions between members of two groups, with within-group distinctions by gender. Call the two groups authors and referees. Their members interact on what they perceive to be a one-to-one basis, with authors seeking a judgment of their work, referees either giving their approval or not. Denote authors as A_i (i = 1,…, I) and referees as R_j (j = 1,…, J). Denote the gender of an author or referee by $f(A_i)$ or $f(R_j)$, equaling 1 for females, 0 for males. The utility of referee R_j when matched with author A_i is:

$$(1) \qquad U(A_i\,, R_j) = \mu_i + \psi_j + \alpha f(A_i) + \beta f(R_j) + \lambda f(A_i)f(R_j) + \varepsilon_{ij}\,.$$

The terms μ_i and ψ_j denote idiosyncratic values of author i's work and referee j's valuation of papers, respectively. Finally, there is some randomly distributed unobservable effect, ε_{ij}, that results with each author-referee match.

A paper is recommended for acceptance if:

$$(2) \qquad U(A_i\,, R_j) > 0\,,$$

which occurs with probability:

$$(3) \qquad \Pr\{\mu_i + \psi_j + \alpha f(A_i) + \beta f(R_j) + \lambda f(A_i)f(R_j) + \varepsilon_{ij} > 0\}.$$

To understand the meaning of the parameters in (3), view the idiosyncratic μ_i and ψ_j as random draws, so that the composite error term is $e_{ij} = \mu_i + \psi_j + \varepsilon_{ij}$. Letting G(.) denote the c.d.f. of $-e_{ij}$, the effect of female authorship on a male referee's acceptance probability is $G(\alpha)$-$G(0)$ and on a female referee's acceptance probability is $G(\alpha+\beta+\lambda)$-$G(\beta)$. The relevant "difference-in-difference" effect is then $[G(\alpha+\beta+\lambda)$-$G(\beta)]$-$[G(\alpha)$-$G(0)]$. This difference-in-difference will be positive (negative) if female referees are comparatively more (less) generous than male referees when matched with female authors.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

The formulation of utility in (1) implicitly assumes that the referees j can identify the gender of authors i. Generalizing (1) and (3) to account for the possibility that the identity of authors is known only in some cases, we obtain:

$$(3') \quad \Pr\{\mu_i + \psi_j + \alpha f(A_i) + \beta f(R_j) + \lambda f(A_i)f(R_j) + \alpha' Z_{ij} f(A_i) + \beta' Z_{ij} f(R_j) + \lambda' Z_{ij} f(A_i)f(R_j) + \varepsilon_{ij} > 0\},$$

where Z_{ij} is an indicator of whether $f(A_i)$ is known to referee j. With this expanded formulation, we would expect $\lambda = 0$ and would infer whether there is within-group favoritism from the difference-in-difference effect, which for $Z_{ij}=1$ (author gender observed) is $[G(\alpha+\beta+\lambda+\alpha'+\beta'+\lambda')-G(\beta+\beta')]-[G(\alpha+\alpha')-G(0)]$.[1]

III. Matching Data to the Model

In order to estimate the parameters describing charity and favoritism, we need a panel of referees and authors that is sufficiently long that individual idiosyncrasies can be accounted for through multiple observations on the same referee matched to different authors. Referees should also be aware of the author's gender, so that there is scope for them to indulge their preferences, if any, for their own gender. Obviously, the latter problem does not arise in laboratory work—parties' gender can, if the experimenter desires, be identified to others. The former problem generally cannot, however, be handled extensively in what are typically very short-duration laboratory experiments.

Our data set contains all the submissions to a leading field journal that were sent out to referees between 1986 and early 2008. Confidentiality restrictions mean we have only the first names of all referees, and most authors, but no information on last names.

[1] A linear specification for G(), as in the linear probability model considered in Section III, leads to simplified estimation of this effect.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

The journal had a strict policy of double-blind refereeing, so that it took some effort for referees to discover whether they matched the author's gender. We cannot know whether or not they knew that a match did or did not exist—whether Z=1 or 0. Evidence from another journal (Blank, 1991), however, suggests that even in the late 1980s referees could identify authors of half the assigned papers. This fact suggests that Z = 1 for many observations.

In case the ease of identifying authors has changed, we need a proxy for Z that might indicate whether it was possible for the referee to make this discovery. Today it is easy to discover the identity of the authors of an unpublished scholarly paper by doing an internet-based search for its title. Such internet discovery was presumably far less prevalent during the early part of our sample.[2] We thus proxy Z by dividing the sample into three periods: 1986-1994, when authors' gender could not be identified via the internet (but perhaps could be through working paper series, direct knowledge of the paper, etc.); 2000-2008, when it could be identified easily via the internet; and 1995-1999, when the degree of identifiability via the internet may have been unclear. In most of our comparisons and estimation, we drop matches from this middle period. To the extent that authors' gender is more likely to be known to the referee (Z=1) in the late period (2000-2008) than in the early period (1986-1994), our ability to detect statistically a gender-matching effect would be greater in the late period.

[2] Whereas 45 percent of American adults in 2000 had reported using the internet in the previous month, only 9 percent had reported doing so in 1996 (*Statistical Abstract of the United States*, various years). A measure of internet access, World Bank, *World Development Indicators*, suggests that only 5 percent of American households even had internet access before 1995.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

2940 initial submissions were sent to at least one referee. In the early period not all authors' given names were listed in the data file, so that, as Table 1 shows, for only 70 percent of the papers were all the authors identifiable. In the later period the records were nearer to being complete. Authors' gender was completely identifiable on 80 percent of the manuscripts, with our inability to identify authors' gender due mostly to ambiguity about the gender identification of various given names. Around one-sixth of the papers for which the gender of all authors could be identified had only female authors, and one-third had at least one female. Both fractions increased between the early and late periods, significantly so for the "any female" category. The regression analysis below will focus on the "any female" classification as the indicator of female authorship ($f(A_i)=1$).

The journal used 6165 referees to judge these papers, and we identified the gender of all but 32. A total of 1514 different referees judged papers during this time period. One instance of refereeing was most common, but 179 individuals judged at least ten manuscripts. Of the identifiable referees, 19 percent were women, a percentage that increased significantly between the early and late periods, as Table 2 demonstrates.

The identification strategy here relies on multiple matches between a referee and a variety of authors, and multiple matches of a particular article to referees. It is thus identical to the strategy used to identify firm-worker match effects by Abowd et al (1999) and, in the context of discrimination, by Parsons et al (2011). This approach is unaffected by the identity of the (very few) editors, so long as they do not assign referees to articles based on their belief that particular referees will or will not discriminate/show favoritism based on the gender of the authors.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

As Table 2 also shows, the matching of authors and referees was not random by gender over the entire period: Female referees were more likely to be matched with papers that had any or all female authors. This was not true during the early period. The relatively few papers that had female authors were only slightly more likely than others to be assigned a female referee; but in the late period there was more gender matching, especially of papers on which all authors were women. The difference in the extent of gender-matching with female-authored papers may reflect the specialization of women in certain sub-fields. Regardless, this phenomenon justifies accounting for this non-randomness in our estimation.

If younger referees are assigned lower-quality papers, and female referees are younger than males, it would be difficult to identify the effects that we seek to measure. It is true that women constituted a growing proportion of the economics profession over this period (Donald and Hamermesh, 2006), so that female referees were probably on average younger (and presumably less experienced) than males. Evidence from a matched sample of referees and authors from what is arguably the leading economics journal (Hamermesh, 1994) suggests, however, almost no quality-matching by reputation of authors and referees.

Each referee was asked to rate the assigned paper on a four-point scale: Accept; accept with minor changes; accept with major changes; reject. As the first two columns of Table 3 indicate, referees recommended that roughly half the papers be rejected. A recommendation of outright acceptance is extremely rare—most positive recommendations involve the referee asking for major changes in the manuscript. The crucial thing to note in the table is the comparison by gender. In the entire sample, and in

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

each sub-period, there is absolutely no evidence of any difference in charitableness by gender: Chi-square tests are very far from rejecting the null hypothesis that the distributions of judgments by male and female referees are the same. Ignoring possible gender differences in the quality of papers that are assigned, the evidence in Table 3 is consistent with the view that the women judging others' work were no more or less charitable than their male counterparts within any time period.[3]

Comparing responses across periods, however, suggests a slightly different implication. While the rejection rate among males rose by 2.1 percentage points between the two periods, the rejection rate among female referees rose by 11.0 percentage points, a double-difference of 8.9 percentage points (s.e. = 3.9). The women who refereed in the late period were significantly less charitable than their male counterparts, as compared to the earlier period.

Figure 1 graphs two-year moving averages of the fraction of papers on which a "no-reject" recommendation was given by female referees and by male referees. The temporal decline appears to be continuous, not something that occurred suddenly during the middle period when the internet became widely accessible. While there is a lot of year-to-year variation, except for the first biennium, one year of which included only four female referees, there is little evidence of discontinuities in these series. There were

[3] The referees' views do matter: All 13 papers rated "accept" by at least one referee and not below "accept with minor revision" by the other(s) were eventually published in the journal. Also, there is substantial, but far from complete agreement on quality by referees. For example, if one referee rated a paper at least "accept with minor revision," 28 percent of the other referees (in two-referee cases) also rated it this highly.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

trends, underlining the desirability of distinguishing between sub-periods, but no permanent sharp changes.

To examine the charitableness issue formally, and to test for gender-matched favoritism, Table 4 presents estimates of model (3'). With very few recommendations of outright acceptance or even acceptance with minor revisions, we define the outcome as non-reject versus reject.[4] Moreover, since the interesting hypotheses concern the interaction variables in model (3'), Table 4 reports results from estimation of a linear probability model (LPM) so that marginal effects are easily inferred.[5] The *Female author* variable is defined to be equal to one if any of the manuscript's authors are female. Results using the "all female authors" classification are extremely similar (see Abrevaya and Hamermesh, 2010) and omitted in the interest of space. The sample is restricted to manuscripts in the early period (1986-1994) and the late period (2000-2008), with the indicator variable *Late* defined to be equal to one for the latter.[6] The reported standard errors are heteroskedasticity-robust and clustered at the referee level.[7]

[4] Estimation that takes advantage of the four-fold classification yields results that are qualitatively identical to those based on the condensed reject/no-reject classification.

[5] A previous version of this paper (Abrevaya and Hamermesh, 2010) reported logit and conditional logit estimates. In this application, the R-squared is particularly low and the predicted probabilities quite close to 0.50, yielding logit marginal effects (and standard errors) that are nearly identical to those obtained from the LPM.

[6] Dividing the entire period into two equal-length parts and replicating the analysis does not qualitatively change any of the conclusions. Also, if we use all observations and replace *Late* by the World Bank measure of internet accessibility, the results are unchanged: There is no evidence of gender favoritism over the entire period, and the triple interaction of the measure of internet accessibility with $f(A_i)f(R_j)$ has a *z*-statistic below 0.5.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

In Table 4, we consider three specifications (columns (1)-(3)) that build up to the full triple-interaction specification in the model in the last of these three columns. Column (1) includes just the *Female author* × *Female referee* interaction; column (2) adds the interactions *Late* × *Female author* and *Late* × *Female referee*; and column (3) adds the triple interaction *Late* × *Female author* × *Female referee*. Columns (4)-(6) present the same three specifications but with referee-specific effects also included, thereby utilizing only within-referee variation in the explanatory variables (so that the *Female referee* coefficient is not identified). All the specifications include a linear time trend (*Year*). While we have no information on referee characteristics (other than gender), we include a quadratic in *Experience*, the number of previous times a person had refereed for the journal.

The results across specifications in Table 4 yield the same conclusions about the possibility of gender favoritism: It is simply not evident in these data. In column (1), where the interaction effect is assumed to be the same in the early and late periods, the coefficient estimate on *Female author* × *Female referee* is small (2.67 percentage points) and statistically insignificant. When interactions with *Late* are introduced in column (2), the coefficient estimate for *Female author* × *Female referee* remains small (3.54 percentage points) and statistically insignificant.

This specification does highlight some trends (early period versus late period) that are unrelated to favoritism/charity. Specifically, as indicated by the *Late* × *Female*

[7] Clustering on referees has little effect on the estimated standard errors, yielding at most five-percent differences from the heteroskedasticity-robust standard errors. Clustering instead at the manuscript level also gives very similar standard errors.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

referee estimate, the referee-gender effect changes significantly from the early to late period. Similar to the raw data, female referees are 2.89 percentage points less likely than male referees to reject a male-authored paper in the early period and 7.12 percentage points more likely in the late period. Also, the significant positive estimate on *Late* × *Female author* indicates a large change in the author-gender effect from the early to late period (female-authored papers 4.48 percentage points more likely to be rejected by a male referee in the early period and 2.29 percentage points less likely in the late period). This change does not necessarily reflect anything about gender favoritism, however, as an increase in the quality of female-authored papers could be the cause.

The complete specification in column (3) yields an estimate of the triple-interaction coefficient very close to zero and statistically insignificant. There is no evidence that the favoritism/discrimination effect changed from the early to late period. Moreover, the point estimates in each period are close to zero, with female referees 4.42 percentage points (*z*-statistic = 0.72) less likely to reject a female-authored paper (as opposed to a male-authored paper) in the early period and 3.04 percentage points (*z*-statistic = 0.64) less likely in the late period. Overall, the results for the separate sub-periods in column (3) show no apparent favoritism/discrimination based on gender matching/non-matching.[8]

[8] One additional possibility is that some of the control variables, particularly experience, are related to the impacts of the gender match and its effects in the separate sub-periods. Interactions of referee experience with the referee-author match and its interaction with the sub-period identifier were essentially zero and had no effect on our conclusions about the possibility of favoritism.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

When we include referee effects in the LPM estimation (Table 4, columns (4)-(6)) and therefore utilize only within-referee variation, the results remain qualitatively very similar (with slightly higher standard errors, as expected). There are some changes in the magnitudes of the female-author effects (i.e., the coefficients on *Female author* and *Late* × *Female author*), but the estimates related to favoritism/discrimination change little from their counterparts in columns (1)-(3). For the complete specification in column (6), we find that female referees are 2.49 percentage points (*z*-statistic = 0.30) less likely to reject a female-authored paper (as opposed to a male-authored paper) in the early period and 4.33 percentage points (*z*-statistic = 0.62) less likely in the late period.[9]

To allow for the possibility that the relevant idiosyncrasies were manuscript- rather than referee-specific, we also estimated a model that included manuscript effects rather than referee effects. Again, the author-referee gender interaction was insignificantly different from zero in both periods. The coefficient estimates on *Female author* × *Female referee* and *Late* × *Female author* × *Female referee* were 0.0178 (s.e. =

[9] Another potentially confounding problem is a gender difference in self-selection. Willingness to complete the assigned task may differ by gender, with women perhaps being more compliant (as they are in their propensity to complete surveys—Moore and Tarnai, 2002). Differential selectivity will only bias the results if the propensity to complete the task is related to differences in the charitableness of doers and refusers. We cannot get at this problem since we have no information on non-compliant referees. If men are more likely not to comply and non-compliers are nastier, then our results are biased in favor of finding that female referees are less charitable than males. A related difficulty may arise from the selection of referees by editors. We cannot solve this problem completely, but the fact that the fraction female referees is slightly above the fraction female in the American Economic Association (Donald and Hamermesh, 2006) should allay some concerns about this issue.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

0.1030) and -0.209 (s.e. = 0.1291), respectively. Finally, when go a further step to account for both referee and manuscript idiosyncrasies, we find no qualitative differences (although standard errors increase, as expected with the increase in the number of fixed effects).

The estimates reported in Table 4 do not account for the possibility, suggested by the across-period comparisons in Table 3, that a change in the mix of referees altered observed behavior between periods. Perhaps women who refereed during the early period were inherently more favorable to female authors, but could not observe authors' gender, while female referees who entered the refereeing pool during the late period discriminated against female authors. This might have occurred because of the increase in female representation in the set of referees—possibly a reduced sense of solidarity among later cohorts of female economists. If this were correct, we would estimate $\lambda' = 0$, even though the agents' preferences exhibited favoritism in one case (the early referees) and discrimination in the other case (the new referees in the late period).

To examine this possibility, we restricted the sample to the 295 individuals who refereed in both the early and late periods. The estimates for this reduced sample are presented in Table 5. Only the complete-specification (triple interaction) results are reported, with columns (1) and (2) of Table 5 corresponding to columns (3) and (6) of Table 4, respectively. A comparison of the results to Table 4 suggests that the apparent lack of favoritism or discrimination was not the result of a change in the mix of referees. Even within this sub-sample there is no evidence of a significant change in behavior toward female authors across the sub-periods, nor of any favoritism overall.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

IV. Conclusions and Implications

Whereas many previous studies have found differences in altruism by gender, our examination of a unique and very large sample on author-referee outcomes in a high-stakes field environment yields no evidence of gender differences. Even accounting for the idiosyncrasies of both the judge and the judged, we still find no such differences. Moreover, there is no evidence of relative favoritism toward one's own gender. The answers to the two sub-titular questions in this study are no.

Female and male economists, at least in this specific setting, appear to behave similarly and in a gender-neutral manner. This might be the result of some inherent sense of fairness, with participants feeling that exercising their prejudices is inappropriate in this particular judging activity—that "their own identity is often tied to their self-concept as experts who are able to stand above their personal interest" (Lamont, 2009, p. 9). Moreover, given the absence of an interaction of experience with gender, our results suggest either that there is no self-selection by gender attitudes, or that fairness/non-discrimination develops very early in the scholars' professional careers. Combined with previous findings, the results imply that gender differences in fairness and favoritism are context-specific. Future research, including laboratory experiments, might examine how the extent of fairness/lack of favoritism depends on the perceived importance of the particular two-sided relationship.[10]

[10] A recent example of a much smaller gender difference in behavior in the real world as compared to laboratory results is Manning and Saidi (2010), although unlike ours that study could not be based on random matching.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

REFERENCES

Abowd, John, Francis Kramarz and David Margolis, "High Wage Workers and High Wage Firms," *Econometrica* 67 (March 1999), 251-333.

Abrevaya, Jason and Daniel S. Hamermesh, "Charity and Favoritism in the Field: Are Female Economists Nicer (to Each Other)?" NBER Working Paper No. 15972, May 2010.

Andreoni, James, and Lise Vesterlund, "Which is the Fair Sex? Gender Differences in Altruism," *Quarterly Journal of Economics* 116 (February 2001), 293-312.

Blank, Rebecca, "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the *American Economic Review*," *American Economic Review* 81 (December 1991), 1041-1067.

Croson, Rachel, and Uri Gneezy, "Gender Differences in Preferences," *Journal of Economic Literature*, 47 (June 2009), 448-474.

Dillingham, Alan, Marianne Ferber and Daniel Hamermesh, "Gender Discrimination by Gender: Voting in a Professional Society," *Industrial and Labor Relations Review* 47 (July 1994), 622-633.

Donald, Stephen, and Daniel Hamermesh, "What Is Discrimination? Gender in the American Economic Association, 1935-2004," *American Economic Review* 96 (September 2006), 1283-1292.

Hamermesh, Daniel, "Facts and Myths about Refereeing," *Journal of Economic Perspectives* 8 (Winter 1994), 153-163.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

Iannaccone, Laurence, "Sacrifice and Stigma: Reducing Free-riding in Cults, Communes, and Other Collectives," *Journal of Political Economy* 100 (April 1992), 271-291.

Lamont, Michèle, *How Professors Think* (Cambridge, MA: Harvard University Press, 2009).

Manning, Alan and Farzad Saidi, "Understanding the Gender Pay Gap: What's Competition Got to Do with It?" *Industrial and Labor Relations Review*, 63 (July 2010), 681-698.

Moore, Danna, and John Tarnai, "Evaluating Nonresponse in Mail Surveys," in Robert Groves, Don Dillman, John Eltinge and Roderick Little (Eds.), *Survey Nonresponse* (New York: Wiley, 2002).

Parsons, Christopher, Johan Sulaeman, Michael Yates and Daniel Hamermesh, "Strike Three: Discrimination, Incentives and Evaluation," *American Economic Review* 101 (2011), forthcoming.

Price, Joseph, and Justin Wolfers, "Racial Discrimination among NBA Referees," *Quarterly Journal of Economics* 125 (2010), forthcoming.

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

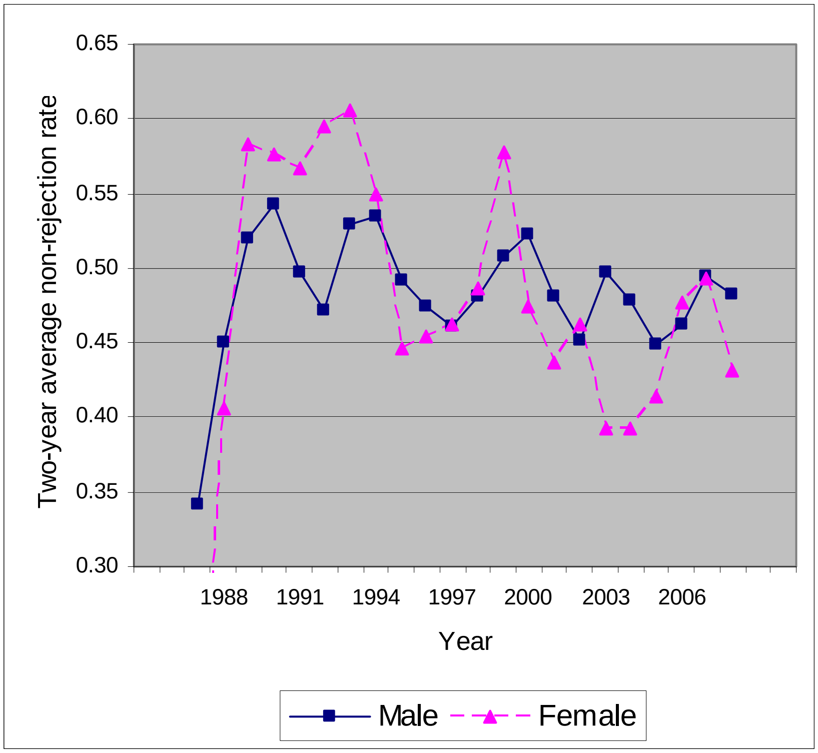

Figure 1. Two-Year Moving Average Rate of Non-Reject Recommendations, By Gender, 1987-2008

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

Table 1. Author Characteristics, 1986-2008 (percentages)

	1986-2008	1986-1994	2000-2008	p-value for test of difference across periods
All authors identified	80.7	69.7	87.3	<0.001
All female authors, if all authors identified	16.9	15.6	18.5	0.104
Any female authors, if all authors identified	35.2	27.8	41.1	<0.001
N =	2940	1116	1095	

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

Table 2. Distribution of Female Referees, 1986-2008

	1986-2008	**1986-1994**	**2000-2008**	**p-value for test of difference across periods**
Percent Female:	18.7	14.3	22.9	<0.001
N =	6133	2347	2173	
Assignment:				
Matched with all female authors, if all authors identified	23.3	16.1	29.0	
Not matched with all female authors, if all authors identified	18.3	14.2	21.8	
p-value (test of random matching)	0.001	0.445	0.004	
Matched with any female authors, if all authors identified	21.8	15.9	25.2	
Not matched with any female authors, if all authors identified	17.7	14.0	21.6	
p-value (test of random matching)	0.001	0.323	0.067	

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

Table 3. Referee Recommendations by Gender, Percentages, 1986-2008

	Full sample 1986-2008		**Early period 1986-1994**		**Late period 2000-2008**	
	Female	**Male**	**Female**	**Male**	**Female**	**Male**
Reject	51.9	51.1	45.8	49.9	56.8	52.0
Major	33.9	34.6	38.7	34.2	31.5	35.4
Minor	13.1	13.1	13.7	14.4	10.9	11.3
Accept	1.1	1.2	1.8	1.5	0.8	1.3
N =	1146	4987	336	2011	498	1675
p-value (test of difference by gender)	0.92		0.41		0.25	

Review of Economics and Statistics Just Accepted MS.
doi:10.1162/REST_a_00163

Table 4: Linear probability model estimates for non-reject probability, full sample (1986-1994, 2000-2008)

	(1)	(2)	(3)	(4)	(5)	(6)
Year	0.0074	0.0073	0.0073	-0.0016	-0.0018	-0.0018
	(0.0032)	(0.0032)	(0.0032)	(0.0062)	(0.0062)	(0.0062)
Late	-0.1215	-0.1224	-0.1230	-0.0948	-0.1110	-0.1102
	(0.0437)	(0.0455)	(0.0456)	(0.0558)	(0.0585)	(0.0587)
Experience	-0.0040	-0.0036	-0.0036	0.0210	0.0219	0.0219
	(0.0043)	(0.0043)	(0.0043)	(0.0087)	(0.0088)	(0.0088)
Experience squared / 100	0.00008	0.00006	0.00006	-0.00048	-0.00052	-0.00052
	(0.00019)	(0.00019)	(0.00019)	(0.00024)	(0.00024)	(0.00024)
Female referee	-0.0249	0.0289	0.0268			
	(0.0295)	(0.0370)	(0.0398)			
Late × Female referee		-0.1001	-0.0961		-0.0901	-0.0953
		(0.0425)	(0.0494)		(0.0642)	(0.0717)
Female author	-0.0014	-0.0435	-0.0448	-0.0256	-0.0817	-0.0803
	(0.0189)	(0.0264)	(0.0281)	(0.0237)	(0.0320)	(0.0336)
Late × Female author		0.0722	0.0747		0.1138	0.1107
		(0.0327)	(0.0369)		(0.0424)	(0.0470)
Female author × Female referee	0.0267	0.0354	0.0442	0.0408	0.0356	0.0249
	(0.0377)	(0.0375)	(0.0618)	(0.0528)	(0.0530)	(0.0819)
Late × Female author × Female referee			-0.0137			0.0184
			(0.0784)			(0.1076)
Referee effects	No	No	No	Yes	Yes	Yes
R-squared	0.0032	0.0056	0.0056	0.1759	0.1784	0.1784
N	4,520	4,520	4,520	3,389	3,389	3,389

Notes: Heteroskedasticity-robust standard errors, clustered at the referee level, are reported in parentheses. The sample size for the referee-effects specifications (columns (4)-(6)) includes only those referees having variation in their non-reject recommendations (i.e., not all zeroes or ones for the dependent variable).

Table 5: Linear probability model estimates for non-reject probability, sub-sample of two-period referees

	(1)	(2)
Year	0.0007	-0.0003
	(0.0050)	(0.0066)
Late	-0.0818	-0.1183
	(0.0609)	(0.0657)
Experience	0.0157	0.0198
	(0.0069)	(0.0094)
Experience squared / 100	-0.00051	-0.00046
	(0.00024)	(0.00025)
Female referee	0.0780	
	(0.0520)	
Late × Female referee	-0.1891	-0.1121
	(0.0661)	(0.0720)
Female author	-0.0406	-0.0753
	(0.0378)	(0.0406)
Late × Female author	0.0677	0.1048
	(0.0479)	(0.0557)
Female author × Female referee	-0.0038	-0.0043
	(0.0788)	(0.0932)
Late × Female author × Female referee	0.0643	0.0782
	(0.1139)	(0.1274)
Referee effects	No	Yes
R-squared	0.0089	0.1877
N	2,277	2,102

Notes: The samples include only referees who reviewed in both the early (1986-1994) and late (2000-2008) periods. Heteroskedasticity-robust standard errors, clustered at the referee level, are reported in parentheses. "Female author" is equal to one if any of the paper's authors is female. The sample size for the referee-effects specification in column (2) includes only those referees having variation in their non-reject recommendations (i.e., not all zeroes or ones for the dependent variable).