Are Women Economists at a Disadvantage in Publishing Journal Articles? A Methodological Comment

WILLIAM F. LOTT*

The findings of a recent article in this journal by Ferber and Teiman (1980) are subject to question in light of their own data. The authors endeavor to conclude from their data that women are discriminated against in the publishing of professional journal articles and that this hinders their advancement. In this comment, we expand the statistical testing methodology beyond that of Ferber and Teiman to search for sex discrimination. As a result of these tests, we suggest that Ferber and Teiman’s results are the consequence of joint authorship and not discrimination.

1 Acceptance Rate

In this section, various statistical techniques are utilized to determine whether there is a difference in acceptance rates between various types of authorship of journal articles. The acceptance rate in general, as noted by Ferber and Teiman, was higher for the sample of double-blind refereed journals than for the non-double-blind refereed journals. We begin our analysis by examining the independence of acceptance rates by authorship type within a given refereeing type. In this initial analysis, we classify articles as "accepted" or "not accepted"; "not accepted" includes both rejected articles and articles being revised but not yet accepted. Table 1 is a condensation of some of the data from Ferber and Teiman’s (1980, p. 191) Table 2. It classifies by authorship for non-double-blind refereed journals all submissions as "accepted" or "not accepted." A chi-squared test for independence of authorship and acceptance rate was performed on the data in Table 1, and we are unable to reject the null hypothesis of independence of acceptance rate and authorship at any meaningful level of significance.1 This is consistent with Ferber and Teiman’s finding from their test of difference of acceptance rates (more will be said about their test procedure later).

Table 2 contains the remaining data from Ferber and Teiman’s Table 2 for double-blind refereed journals with respect to acceptance versus nonacceptance and authorship of submitted articles. A chi-squared test for independence of acceptance rates and authorship for Table 2 produces a chi-square which is significant at the 5% level of significance. This suggests that there is some apparent difference in acceptance rates between types of authorship within the group of articles submitted to double-blind refereed journals.

*Associate Professor, University of Connecticut, Storrs, Connecticut, 06268. This comment has benefited from the helpful suggestions of Polly Allen, Arthur Goldsmith, Steven Miller–colleagues at the University of Connecticut, and Ms. Tina Lawrie–a Ph.D. recipient from the University of Connecticut. In conformity with tradition, I accept responsibility for errors remaining.

1The critical values of a chi-square with 2 degrees of freedom for the 5%, 1%, and 0.1% levels of significance are 3.84, 6.63, and 10.649 respectively.
sults of the Tukey-Cramer test clearly indicate that if the proportion of articles submitted by the 8th group that were accepted, 1 if articles in 8th group were blind refereed, 0 if articles in group were not blind refereed, 1 if articles in group were accepted, 0 if articles in group were not accepted, 1 if articles in group were not blind refereed, and 0 if articles in group were blind refereed. The basic model used is given in equation 1.

\[ \ln(P_i / 1 - P_i) = a_0 + a_1B_i + a_2M_i + a_3F_i + u_i \]  

A multiple-comparisons test for double-blind refereed journals. Some of this information is given for comparison sake only. One would not normally carry out the test on the non-double-blind refereed journals given that the chi-squared test on Table 1 was not significant. We have modified the traditional studentized range test to make our test statistic more comparable to the Z of the normal test. The traditional studentized range statistic would have been

\[ u = P - P^* / \sqrt{(1/m_n + 1/m)} \]  

where \( x \) is the square root of the estimated variance of the Bernoulli random variable \( P \) which is defined as the proportion of articles submitted by the \( i \)th group that were accepted, and \( P^* \) equals the overall proportion of accepted articles. In our modification we left \( x \) out of the numerator. Hence in every case one test we had to transform the critical value of the studentized range \( u \) into the critical value of our test statistic \( z \) by dividing \( u \) by \( \sqrt{n} \), the square root of \( x \).

The studentized range is a form of \( z \). In this test, if one asks if an observed larger sample proportion is in fact statistically larger than the smaller observed sample proportion, this is a one tail test. Therefore, to make appropriate comparisons we set the test up as a one-tail test. Since in reality the test is an \( z \)-test, we have done it in fact equivalent to a 2-tail test for the difference of proportions at the 10% level of significance.

Although Ferber and Teiman’s Table 2 contains 12 journals and 3 authorship classifications, we were able to only obtain 35 observations since there were no submissions in the mixed classification to the EASTERN ECONOMIC JOURNAL. Upon fitting our logit model using OLS, we obtained the following results:

\[ \ln(P_i / 1 - P_i) = -1.64 + 0.52B_i + 0.24 \]  

Where \( x = 0.24 \) and \( x = 0.30 \), respectively. In our modification we left \( x \) our test we had to transform the critical value of the studentized range \( u \) into the critical value of our test statistic \( z \) by dividing \( u \) by \( \sqrt{n} \), the square root of \( x \).

The studentized range is a form of \( z \). In this test, if one asks if an observed larger sample proportion is in fact statistically larger than the smaller observed sample proportion, this is a one tail test. Therefore, to make appropriate comparisons we set the test up as a one-tail test. Since in reality the test is an \( z \)-test, we have done it in fact equivalent to a 2-tail test for the difference of proportions at the 10% level of significance.

Our equation furnishes us with three pieces of information. The least interesting, since it was expected, was that one has better odds of being published in the set of blind refereed journals than the set of non-blind refereed journals. The next two conclusions are more interesting. First, that there is no significant
difference in the odds of being published between strictly male authored articles and strictly female authored articles. Second, that mixed authored articles have better odds of publication that articles of single sex authorship.

Before drawing conclusions from the above results, we will give Ferber and Teiman’s hypothesis one more test (a test similar to their approach). For this test, a second logit model is constructed as follows:

\[
\ln \left( \frac{P}{1-P} \right) = C_0 + C_2 R + C_3 A F + U,
\]

(2)

Where, \( AF = \) 0 if authorship was strictly male, and 1 if authorship had at least one female, and all other variables are as defined above. The estimated model yields

\[
\begin{align*}
\ln \left( \frac{P}{1-P} \right) &= -1.60^{**} + 0.50^{*} R \\
&\quad + 0.26^{*} AF \\
&\quad + 0.13 F \chi^2 = 2.44^{*}
\end{align*}
\]

** = significant at 25% level of significance,

* = significant at 10% level of significance, and

*** = significant at 1% level of significance.

It is difficult to argue the existence of a regression for the above model. If one does, however, one cannot find any support for discrimination against articles that have at least one female author. In fact, though the coefficient is insignificant, its sign suggests a positive effect for articles authored by at least one female when we control for type of refereeing. Clearly, these results do not support Ferber and Teiman’s conclusions about discrimination.

II Submission Decision

A possible explanation of the higher acceptance rate of articles that are exclusively or jointly authored by women in blind refereed journals is that the better and more prolific women authors perceive discrimination in refereeing and hence endeavor to avoid it by submitting their articles to blind refereed journals. If this hypothesis is true then the sex of the author or coauthor of a submitted article should not be independent of the type of refereeing (blind versus non-blind) of the article. Table 4 contains a condensation of the information in Ferber and Teiman’s Table 2. On applying a chi-squared test of independence to Table 4, one can reject the null hypothesis of independence of sex of author or coauthor of a submitted article should not be independent of the type of refereeing (blind versus non-blind) at the 25% level of significance (2 degrees of freedom, \(X^2 = 1.44\) and \(X^2(2; df = 2.77)\)). Upon condensing the table as Ferber and Teiman did to just two classifications of authorship (male only and at least one female), however, one can no longer reject at any meaningful level of significance the null hypothesis of independence in submission of authorship and refereeing type (\(X^2 = 0.44\) and \(X^2(1; df = 1.32)\)).

Given the above findings and Ferber and Teiman’s findings, one suspects that there is some inherent difference in the situations of articles authored by females only and those authored by a combination of at least one female and at least one male. This belief is further supported if one does a test of difference of proportions on the submission rates of particular groups between types of refereeing. Table 5 contains the proportions that are of interest to us. Of the three groupings considered in Table 5 only mixed authorship shows a significant difference at the 10% level of significance in the rate of submission to blind versus non-blind refereed journals. Clearly, mixed authorship appears to be the factor that is causing the difference that we are observing in submission rates as well as the cause of the difference observed by Ferber and Teiman in acceptance rates.

III Conclusions

Given that we cannot support Ferber and Teiman’s results of discrimination, what do our results suggest? All the findings of sections I and II indicate that the critical factor in all our analysis is mixed authorship. For the moment ruling out any argument along the line of chemistry between the sexes, we conclude that the critical factor is joint authorship. The classifications male only and female only contain both singly and jointly authored articles. The classification mixed authorship clearly implies jointly authored articles. Thus, it appears that jointly authored articles have higher probabilities of acceptance. What we may be observing in a case of division of labor or specialization between authors that leads to an overall higher-quality product and hence, a higher acceptance rate. A high-quality, publishable article can be produced by joint authors with less total ef-

### Table 4

<table>
<thead>
<tr>
<th>Type of Referring</th>
<th>Male only</th>
<th>Mixed</th>
<th>Female only</th>
<th>Total Submissions by Referring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-blind</td>
<td>5627</td>
<td>191</td>
<td>186</td>
<td>5914</td>
</tr>
<tr>
<td>Blind</td>
<td>3664</td>
<td>84</td>
<td>105</td>
<td>3843</td>
</tr>
<tr>
<td>Total Submissions by Authorship</td>
<td>9281</td>
<td>185</td>
<td>291</td>
<td>9759</td>
</tr>
</tbody>
</table>

### Table 5

<table>
<thead>
<tr>
<th>Type of Referring</th>
<th>Female only</th>
<th>Mixed</th>
<th>At Least One Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-blind</td>
<td>2.9%</td>
<td>1.7%</td>
<td>4.6%</td>
</tr>
<tr>
<td>Blind</td>
<td>2.7%</td>
<td>2.2%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

* z for testing difference of proportions 0.59, 1.72, 0.00
fort than it would take a single author to reach the same level of quality.

Finally, if our promotion and tenure systems discount jointly authored articles by less than 50% then it would pay joint authors to expend a combined effort in producing an article greater than the effort of a single author. Hence it is possibly both the reward structure and specialization that has lead to the phenomenon recorded in Ferber and Teiman’s data that mixed authored articles have better odds of publication. It is a joint authorship effect and not a discrimination effect that Ferber and Teiman’s data displays.

IV References


Reply to Professor Lott

MARIANNE FERBER

Lott raises several objections to the methods used, and the conclusions drawn by Ferber and Teiman (1980). F and T henceforth.

He also suggests alternative approaches and different interpretations. We shall briefly explain why we do not find either his methods or his conclusions acceptable.

Lott chooses a logit model to test our data. It is customary to use logit when a “yes” or “no” dichotomy is involved. Such an approach would be appropriate if he used the decisions for individual articles, rather than percentages. In that case, however, the N would be far larger, and his conclusions might well be different. He further uses ln’s of the percentages, without even making clear why he considers this preferable, or even acceptable.

The reason for using logit is that Lott believes a series of tests of differences for the data in Table 2, F and T is not appropriate, because the tests are not independent. He is right, of course, that the categories of “at least one woman author” and “mixed authors” are not independent, since the former category includes the latter, but this is not true of the categories “male only,” “female only,” and “mixed authors.” Our findings that the acceptance rate for manuscripts written by only women was higher for journals with double-blind refereeing, just as was true for “mixed authors” stands.

Furthermore, while Lott’s claim that the use of a logit model permits him to test for discrimination in acceptance rates across both types of refereeing simultaneously, it is precisely for this reason that the test is not useful. When women’s acceptance rates are lower in one data set, and higher in the second, putting them into one regression is not a way of testing for the difference, but of obscuring it. The same problem arises with the second proposed logit model. Furthermore one may wonder how serious Lott is about this one, since he himself says that it is difficult to argue the existence of a regression for it.

Let us next turn to Lott’s claim that odds of being published between strictly male authored articles and strictly female authored articles are not significantly different. The variable in his regression, on which this conclusion is supposedly based, is defined $F = 0$ if authorship of i-th group was not wholly female; $F = 1$ if i-th group wholly female. He is actually comparing strictly female authored articles with those that are strictly male authored, plus those that have mixed authorship. Thus Lott tests the difference between two groups never actually compared in F & T.

Last we turn to the suggested interpretation for the agreed upon fact that “mixed” authors have a particularly high acceptance rate. It is very puzzling that Lott, having just established to his satisfaction that there is no discrimination against female authors, would now argue that the better and more prolific ones “perceive discrimination in refereeing and hence endeavor to avoid it,” and that sex of author(s) is therefore not independent of refereeing. Given the tenous nature of this reasoning, it is not surprising to find little empirical support for this hypothesis. Of the three groupings considered only one shows a significant difference, and that only at the 10% level, between journals with and without double-blind refereeing. Nor is it easy to see why “mixed authors” but not women should