Seminar Paper No. 730

GENDER AND OVERCONFIDENCE

by

Claes Bengtsson, Mats Persson and Peter Willenhag

INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES
Stockholm University

Seminar Paper No. 730

Gender and Overconfidence

by

Claes Bengtsson, Mats Persson and Peter Willenhag

Papers in the seminar series are published on the internet
in Adobe Acrobat (PDF) format.
Download from http://www.iies.su.se/

Seminar Papers are preliminary material circulated to
stimulate discussion and critical comment.

June 2004
Institute for International Economic Studies
S-106 91 Stockholm
Sweden

Gender and Overconfidence

by

Claes Bengtsson[a], Mats Persson[b*], Peter Willenhag[c]

Abstract:

Do males differ from females in terms of self-confidence? The structure of the Economics I exam at Stockholm University provides an opportunity to shed some light on this question. By answering an extra, optional question, the students can aim for a higher mark. We find a clear gender difference in that male students are more inclined than female students to take this opportunity. This difference in self-assessment is more pronounced among younger than among older students.

Keywords: Overconfidence, gender differences, exam behavior, exam results.

JEL classification: A2, J2, J16.

[a] Department of Economics, Stockholm University, SE-106 91 Stockholm, Sweden
[b] Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden
[*] Corresponding author. Tel. +46 8 162224. Fax +46 8 161443.
E-mail address: mp@iies.su.se (M. Persson).
[c] Department of Economics, Stockholm University, SE-106 91 Stockholm, Sweden

Gender and Overconfidence

1. Introduction

There is a small but growing literature indicating that men are more overconfident than women. For instance, Barber and Odean (2001) show that male investors trade more than female investors, presumably due to higher confidence in their own ability.[1] While highly relevant to our understanding of social phenomena, studies of the link between gender and overconfidence is hampered by a lack of good data. The purpose of the present paper is to exploit a new data source to shed light on the issue.

For several years, Stockholm University has used a particular design of the written exam for the first-year courses in economics. There are only three grades: Very Good (VG), Pass (P), and Fail (F). The exam consists of four questions, and in order to get the grade P on the entire exam, the student needs a P on each one of these four questions. For the student who aims for a VG on the exam, however, there is a fifth question. In order to get such a high grade, the student needs firstly a VG on each of the first four questions and secondly a satisfactory answer to Question 5. If one has a mere P on one or more of the first four questions, one can never get a VG on the exam, regardless of whether or not one gives a good answer to the fifth question. Of course, at the time when one decides whether or not to answer Question 5, one does not yet know how good the answers to Questions 1-4 are. This is where self-assessment enters the picture. Furthermore, if one is satisfied with a mere P on the exam, one has no incentive to answer the fifth question – that would only be a waste of time.

This exam structure provides an interesting opportunity for studying gender differences in self-assessment among the students. The issue we analyze is whether female students are less prone to answer the fifth question than are male students. For this purpose, we use historical exam data consisting of enough observations so that reliable significance tests can be obtained.

[1] See also Correll (2001).

We find, in fact, that there is such a difference in exam behavior between male and female students; men tend to be more inclined to answer Question 5 than women. But this is not the entire story. There are interesting patterns within each gender group. Dividing the students into "good" and "mediocre" ones (depending on their results on Questions 1-4), we find that the gender difference in self-assessment was more pronounced in the mediocre group. Furthermore, we find an age effect: the gender difference in self-assessment is limited to the younger students.

2. The Empirical Picture

2.1. The Data

We have looked at the results on the Microeconomics I exam, which is the first exam in economics that the students take. Data has to be recorded and preserved in such a form that we can see how each student has responded to each separate question. We have used five Microeconomics I exams, from the Fall Term 2001 through the Spring Term 2004, that satisfy this informational requirement. In total, there were 2217 students who took those exams, of which 1102 (49.7%) were female and 1115 (50.3%) were male.

2.2. Exam Results

Among the female students, 78.8% (869 out of 1102) passed the exam, which means that they received the grade P or VG. Among the male students, 76.5% (853 out of 1115) passed the exam. Is 78.8% significantly different from 76.5%?

To answer this question, we use the test statistic

$$t = \frac{\hat{\mu}_F - \hat{\mu}_M}{\sqrt{\frac{\hat{\sigma}_F^2}{n_F - 1} + \frac{\hat{\sigma}_M^2}{n_M - 1}}},$$

where $\hat{\mu}_j \equiv n_j^i / n_j$, $j = F, M$, are the sample fractions among female and male students, respectively, who passed the exam, and where $\hat{\sigma}_j^2 \equiv \hat{\mu}_j(1 - \hat{\mu}_j)$, $j = F, M$,

are the sample variances. This statistic is *t* distributed with $n_F + n_M - 2$ degrees of freedom.[2]

The null hypothesis is now that the percentage of female student who pass the exam is the same as the percentage of male students who pass the exam. Substituting $n_F = 1102, n_F^i = 869, n_M = 1115$ and $n_M^i = 853$, where "*i*" indicates "passed the exam", into the equation above, we obtain *t* = 1.3312, which is significant at the 90 percent confidence level. Thus the observed frequencies 78.8% and 76.5% are sufficiently different for us to be able to reject the null hypothesis that male students pass the exam as often as female students do. This result is in line with earlier evidence that females seem to perform slightly better than males at school.[3]

What about the VG grade? There were 130 (11.8%) female students with a VG, and 180 (16.1%) male students; the difference in proportions achieving a VG are statistically significant at the 99 percent level (*t* = -2.9574).

There are thus striking differences between male and female students in terms of outcome: female students are slightly better at passing the exam, but male students are much better at getting the highest grade. Whether this pattern is due to differences in innate intellectual capacity, study habits, or some other factor, is an open question that deserves further investigation. In this paper, we confine the analysis to investigating whether there are any gender differences in exam behavior.

2.3. Exam Behavior

Among the female students, there were 506 who were qualified to answer Question 5, i.e. who scored a VG on each of the Questions 1-4. Of these, 424 (83.8%) answered Question 5. Among the male students, 480 were qualified, of whom 418 (87.1%) answered Question 5. Applying the above test statistic, we find that qualified females are less prone to answer Question 5 than qualified males. The difference is significant at the 90 percent level (*t* = -1.4653).

[2] See, for instance, Hogg and Tanis (1997, pp. 305-309).
[3] Cf. McNabb, Pal and Sloane (2002).

There were 596 unqualified females (i.e. those who had not received a VG on each one of the Questions 1-4). Of these, 248 (41.6%) still tried to answer Question 5. Among the 635 unqualified males, 309 (48.6%) were similarly optimistic. Here the sample proportions are statistically significant at the 99 percent level (t = -2.4891). We can thus reject the null hypothesis that females are as prone as males to go for Question 5. This pattern holds for both qualified and unqualified students, but it is more pronounced among the unqualified ones.

There are other patterns in the data. For instance, we divided the unqualified students into two groups: those who are "mediocre" in the sense of having at least the grade P, but not the grade VG, on the Questions 1-4, and those who are "bad" in the sense of having an F on at least one of the Questions 1-4 (i.e. failing the exam). In both groups, males were significantly more prone to answer Question 5; the significance level was 99 percent (t = -2.4153) in the "mediocre" group, while it was 90 percent (t = -1.4401) in the "bad" group.[4]

Finally, we investigated whether there was an age effect involved. We divided the students into two age groups of roughly equal size: those between the ages 18 and 22 (inclusive), and those between 26 and 62 (inclusive). We then tested the following null hypothesis: Within each age group, the propensity to answer Question 5 is the same for men and women.

Running the same tests as above for the younger group of qualified men and women, we obtained t = -1.0280, i.e., there was no significant difference in behavior between younger qualified men and women. Running the test for the younger group of unqualified men and women, we obtained t = -1.5317, i.e., there was a significant difference (at the 90 percent level) between younger unqualified men and women. Running the test for the older group, we did not, however, observe any significant difference between men and women. Breaking the data down into smaller and smaller subgroups, however, may make it difficult to discern significant differences due to the correspondingly fewer degrees of freedom. If there is a significant difference in the

[4] There seems to be a teacher-specific effect involved here; the patterns are slightly different for different teachers over the years, but the general pattern is nevertheless robust.

entire data set, there may still be no significant difference in any of the subgroups, provided they are sufficiently small.

Summing up, there is a clear difference in self-assessment between men and women. With the qualification for the problem of degrees of freedom, this difference seems to be more pronounced in the younger age group.[5]

3. Some Explanations

There are clear gender differences both in exam results and exam behavior. Although women were slightly better than men in the sense of passing the exam, men were significantly better in terms of obtaining a VG. This is hardly surprising, since we have shown a higher propensity among men to aim for the higher grade. An economist would attribute this difference in behavior to constraints, preferences, and/or technology.

3.1 Constraints

The most immediate constraint is the budget constraint, which in this case is a time constraint: answering the optional fifth question takes time. If women's time were more valuable than men's time, it would be rational for the former not to waste time on Question 5. This could be tested. At the 2003 Spring Term exam, we collected data on the time at which each student handed in his or her answers. It turned out that there was no significant difference between men and women in how much time an individual spent answering the questions. This holds for the entire population (380 students) as well as for the sub-sample of persons who chose to answer Question 5 and for the sub-sample of persons who did not.

Another aspect of the time budget constraint could be that a student might be inclined to skip Question 5 in order to have more time for Questions 1-4. To test this, we divide all students into two groups: those who answered all five questions, and those who answered only four of them. Using the same test statistics as before, it turned out

[5] Another interesting question is whether there is an age effect within each gender group. It turned out that this is the case at the 99 percent level. In our sample, younger women were more prone to answer Question 5 than older women ($t = 2.52$), and younger men were much more prone to answer Question 5 than older men ($t = 3.04$).

that those who answered all five questions spent on average 30 minutes more on the exam than those who did not answer all five questions. This applies to the whole sample as well as to males and females separately. While the time difference of half an hour is statistically significant (some students may have been more in a hurry than others) there is no gender difference in this respect.

Thus there is no indication that female students regard their time as more valuable than do male students, or that they make some other type of time substitution across the exam questions.

Having dismissed the time budget constraint as a possible explanation to the gender difference between male and female student exam behavior, we cannot rule out that there are other, more subtle constraints. For instance, North (1990) refers to formal and informal institutional constraints. The latter, i.e., social norms and traditional gender roles, are difficult to measure but may nevertheless be binding for, in particular, women with a low degree of self-esteem.

3.2 Preferences

The distinction between preferences and informal institutional constraints is of course ambiguous. A difference in preferences that may be the result of social norms could be, for instance, optimising versus satisficing behavior. If men were optimizing, while women were satisficing, this would be consistent with the observed differences in exam behavior. Our data does not, however, allow for a sharper test of this hypothesis.

Another preference trait could be that of risk aversion. This does not, however, provide an explanation to the observed gender differences in our experiment. Question 5 could be regarded as a lottery ticket. If such a lottery ticket were costly, more risk-averse persons would be less inclined to buy it. But if the ticket were offered for free, everybody would take it, regardless of his or her degree of risk aversion. Now, if the time cost of answering Question 5 is negligible, failing to answer the question would be similar to failing to pick up a lottery ticket that is offered for free. Differences in the propensity to pick up such a ticket cannot be explained by differences in risk aversion.

3.3 Technology

In addition to constraints and preferences, there may be differences in technology. One possibility is that unqualified men who choose to answer Question 5 really are marginally better than the females – but this difference is not evident in our coarse grading system. To test this hypothesis, we collected detailed individual scores on Questions 1-4 for two of the exams (Spring Term and Fall Term 2003 with a total of 758 students), and we used these data to address two questions: a) Is the mean score on Questions 1-4 of unqualified men different from that of unqualified women? and b) Is the mean score on Questions 1-4 of unqualified men who chose to answer Question 5 different from that of unqualified women who chose to answer Question 5? For both these questions, the answer turned out to be negative: there is no significant difference in mean score between men and women (in fact, the women had better average scores than the men – although not significantly so). There is thus no evidence that the higher propensity among males to answer Question 5 can be attributed to marginally better scores on Questions 1-4 among the males.

Another possibility that has to do with technology is that men may be more uncertain about the outcome of the exam. Technically speaking, their subjective posterior probability distribution (after having seen the exam questions) of their score might have a higher variance than the corresponding distribution of the women. Note that this is a different argument from the one concerning overconfidence above. You can be quite confident, in the sense of having a high expected value of your exam score, but still think that the outcome (which is, after all, a stochastic variable) has a high variance. Now, if men were more uncertain, it would be a rational strategy for them to take a chance with Question 5, provided that their time is not too valuable.

It is hard to find out the variances of people's subjective probability distributions, but we can at least test whether the actual exam scores of men have a variance different from that of women. To do this, we calculated the variances of the exam results for men and women. For Questions 1-4, an F yields 0 points, a P yields 1 point, and a VG yields 1.5 points. A maximum of 6 points could thus be obtained. Assigning numerical values to qualitative outcomes like this is of course somewhat arbitrary, but it makes it possible to compute estimated variances (we also tried other numerical

values, with similar results). The variance over the results on Questions 1-4 among the males was 1.151 and the variance among females was 0.850. These variances are significantly different at the 99 percent level.[6] A higher dispersion among men than among women could of course have many explanations, but it indicates that it might be rational for men to be more uncertain about their abilities – or at least about the outcome of the Economics I exam.[7]

There are thus a number of potential explanations, in terms of constraints, preferences and technology that are consistent with the data. To discriminate between these explanations would require quite another data set, a data set that is presently not available.

4. Concluding Comments

One major question emerges from our findings: Why do men and women display different behavior? This also raises some other interesting issues, for instance the difference in behavior between younger and older students. The answers to these questions may be found in terms of differences in objectives, in constraints, or in technology. This seems to be a promising area for future research.

There are also two supplementary questions. The first is whether the gender difference in exam behavior (i.e., males are more likely to aim for a VG) can fully or partly explain the gender difference in exam outcome (i.e., males are more likely to obtain a VG). This is not self-evident. The difference in outcome can have other reasons. For instance, if men and women have the same average ability, a higher variance in abilities among men would necessarily lead to both more men failing the exam and more men obtaining the highest grade – just like in our data on the exam results. The only way to find out what is actually driving the higher incidence of the

[6] For the relevant test statistic of the null hypothesis that the ratio of the two variances is equal to unity, see Hogg and Tanis (1997, pp. 312-316). A number of other ways of computing the variance were used. For instance, Question 4 consisted of 10 sub-questions. Assigning one point per correct sub-question answer (i.e., the minimum is 0 and the maximum, 10) yields the estimated variance for men as 6.048 and for women, 6.023. This difference is not statistically significant.

[7] To shed further light on this issue, as well as on the issue of the value of time, one would like to have data on the amount of time spent on preparing for the exam. If females spend more time than males studying for the exam, this would be consistent with women being better at passing the exam, as well as males having a higher uncertainty about their knowledge.

VG score among men is to make a controlled experiment with another exam format that does not build on the students' self-assessment.

The other question is whether or not the present exam format discriminates against women. This depends on the answer to the previous question, but even if the difference in outcome *could* be fully explained by the exam design, this would not necessarily be discrimination. If the labor market demands people who do not hesitate to pick up a free lottery ticket, the present exam format may be seen as a screening device to find out which ones have the abilities demanded by the labor market. This need not be discrimination, and changing the format so that the exam outcome does not depend on the students' propensity to go for a higher grade would thus reduce the information content of the exam results.

Without taking any stand on any of these issues, we look forward to more data becoming available, under different exam formats.

Acknowledgements

We are grateful for comments from Mahmood Arai, Anne Boschini, Lena Nekby, Claes-Henric Siven and Peter Skogman Thoursie. We also want to thank Mustafa Pala for data management.

References

Barber, B. M., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. Quarterly Journal of Economics 116, 261-292.

Correll, S. J., 2001. Gender and the career choice process: The role of biased self-assessments. American Journal of Sociology 106, 1691-1730.

Hogg, R. V., Tanis, E. A., 1997. Probability and Statistical Inference, Fifth edition. Prentice-Hall, Upper Saddle River, N. J.

McNabb, R., Pal, S., Sloane, P., 2002. Gender differences in educational attainment: The case of university students in England and Wales. Economica 69, 481-503.

North, D.C., 1990. Institutions, institutional change and economic performance. Cambridge University Press, Cambridge and New York.

SEMINAR PAPER SERIES

The Series was initiated in 1971. For a complete list of Seminar Papers, please contact the Institute.

2003

717. *Assar Lindbeck*	Improving the Performance of the European Social Model - The Welfare State over the Life Cycle.
718. *Harry Flam*	Turkey and the EU: Politics and Economics of Accession.
719. *Assar Lindbeck*	An Essay on Welfare State Dynamics.
720. *Gino A. Gancia*	Globalization, Divergence and Stagnation.
721. *Dirk Niepelt*	Tax Evasion Dynamics.
722. *Dirk Niepelt*	Intra-Generational Conflict: The Role of Balanced Budget Rules.
723. *Peter Svedberg and John E. Tilton*	The *Real* Real Price of Nonrenewable Resources: Copper 1870-2000.
724. *Peter Svedberg*	World Income Distribution: Which Way?
725. *Assar Lindbeck and Dennis Snower*	The Firm as a Pool of Factor Complementarities
726. *Davide Ticchi and Andrea Vindigni*	Endogenous Constitutions

2004

727. *Conny Olovsson*	Why do Europeans Work so Little?
728. *Conny Olovsson*	The Welfare Gains of Improving Risk Sharing in Social Security
729. *Conny Olovsson*	Social Security and the Equity Premium Puzzle
730. *Claes Bengtsson, Mats Persson and Peter Willenhag*	Gender and Overconfidence

ISSN 0347-8769
Stockholm, 2004
Institute for International Economic Studies