Is Team Formation Gender Neutral? Evidence from coauthorship patterns.*

Anne Boschini
Stockholm University

Anna Sjögren
The Research Institute of Industrial Economics

August 9, 2004

Abstract

We model the formation of teams as a random matching process influenced by the agents' preferences for team size and gender composition. We test hypotheses regarding gender and team preferences on the pattern of coauthorship in articles published 1991-2002 in three top economics journals. We find that the female/male gap in the probability of having a female coauthor increases with the proportion of female authorships in the field. This, together with the finding that women single author significantly more than men and that female single authorship declines more than male ditto as the share of women increases, allows us to reject gender neutrality in team formation in favor of an hypothesis stating that the fraction of individuals who prefer teaming up with their own sex is larger than the fraction who prefer the opposite sex.

Keywords*:* Team formation, Gender Preference, Segregation, Coauthorship patterns.

JEL: A14, J16, J41, M50.

*We thank seminar participants at Stockhom University, SOFI and IUI for useful comments. Assistance with gender coding from Jakob Eliasson, Christina Håkansson, and Gulay Özcar is gratefully recognized.

1 Introduction

Teamwork is a feature of many professional activities. Teamwork makes specialization possible and will hence enhance productivity. Becker and Murphy (1992) argue that the realization of these potential returns from specialization will depend on the ability to coordinate the efforts of the team members. When coordination costs are too high teams are not formed. Both gains and costs from teamwork are linked to the heterogeneity of the team members. One potentially important dimension of heterogeneity is gender. The question posed in this paper is how important gender is in team formation? More specifically, do agents who form teams favour segregation, integration or are they neutral? We address this question by developing a simple model of team formation in the presence of team size and gender preferences. We then test our theoretical predictions on the coauthorship pattern in articles published in the American Economic Review, the Journal of Political Economy and the Quarterly Journal of Economics between 1991 and 2002 and find that our data is consistent with a preference for gender segregation on the part of male or female researchers, or both.

This evidence of preferences over the gender composition of teams, whether taste or coordination cost driven, is potentially important for understanding gender differences in career choice and persistence in occupational gender segregation.[1] Clearly, when teamwork is an important determinant of individual productivity, the availability of productive team mates, who are willing to collaborate is going to enter the career decisions of individuals. Anyone considering the hard work required to reach top management positions or to become a professor in Economics needs to take into account the prospects of being invited and accepted on a board of directors or to find good collaboration partners. Whenever there is a preference for gender segregation in team formation, these prospects will be smaller for the gender in minority.

Academia, and economics in particular, is a profession where teams are formed voluntarily. As a result, the teams that are formed presumably reflect the preferences and perceptions of returns to specialization and costs of coordination of the team members. Economics is also a very male dominated profession. This, together with the ease at which output produced by the teams can be measured and quality controlled for, i.e. publication in academic journals, makes the collaboration pat-

[1] See e.g. Breen and Garcia-Penalosa (2002) for references on occupational gender segregation.

tern of economists a suitable candidate for the study of gender preferences in team formation. Moreover, while some aspects of the return to specialization and the coordination costs are specific to the field of economics, and hence perhaps of little interest outside academia, there is less reason to believe that the gender pattern in team formation in economics is specific only to the economics profession. It is arguably the case that gender preferences revealed by the gender pattern in team formation in economics are indicative of a presence of gender preferences also in other professions with similar degrees of complementarity between team mates and comparable sex ratios.[2]

We first develop a simple random matching framework where the researchers in a field voluntarily form teams in order to maximize the value of output i.e. the utility they derive from research. We assume that researchers may differ in their valuations of team size and gender of team mates but that the distributions of team size preferences and gender preferences are independent. As a result, some researchers prefer writing alone, some prefer team work and some prefer either or depending on the gender of the prospective team mate. Importantly, we allow the preference distributions to be gender specific. Since our interest is gender preferences, we abstract from coauthorship decisions stemming from productivity differences across authors, which is a driving force in Goyal et al (2004). We then derive a set of testable predictions regarding how the share of women in the pool of possible collaborators affects the male and female propensities to work alone (single author) and form teams with either sex, given a set of different assumptions regarding the male and female distributions of team size and gender preferences.

We confront our theoretical implications with data on economics publications. We gender code the authors of all articles published in the American Economic Review, the Journal of Political Economy and the Quarterly Journal of Economics between 1991 and 2002. We can confirm an increasing trend in coauthorships observed by others. Interestingly, we find that the decline in single authorships is sharper for women than for men, possibly reflecting that more women now stay in academia beyond publishing their single authored job market paper. Towards the latter half of our period of study, the behavior of men and women have converged

[2]In the light of results on homogenous vs heterogenous teams in Prat (2002), we should add that a possible preference for gender homogenous teams in economics could reflect strong complementarities in team production that makes heterogeneity among team mates disadvantageous.

regards the choice between coauthorship or single authorship. Women no longer appear to be at a disadvantage when it comes to finding coauthors, as was found by McDowell and Kiholm Smith (1992) for the 1980's.

Following Laband and Tollison (2000), we make use of the articles' JEL-code classifications. This allows for field specific costs and returns to team work and it also permits us to construct an article specific measure of the gender ratio in the relevant pool of prospective collaboration partners. However, controlling for JEL-codes does not resurrect the idea of gender neutrality in coauthorship patterns. Using binomial logit (and probit) analysis, we find that the more balanced the ratio of women to men within a field, the more gender segregated the choice of coauthors. While the propensity of men to coauthor with a woman is rather insensitive to the share of women publishing in the field, the female propensity to coauthor with a woman increases with the share of women in the field. Such a pattern could, in fact, be consistent with gender neutrality if men were sufficiently more interested in single authorships than women. Given that our data tell us that men are significantly less likely to single author than women and, moreover, that the female propensity to single author is more sensitive to the female fraction of publications within a field, we can reject the hypothesis of gender neutral preferences. Instead, the pattern found in the data is consistent with a preference for gender segregation on the part of male or female researchers, or both.

Our analysis takes the gender ratio, as well as the distribution of preferences, in the pool of collaborators as given. This is, of course, a simplification since these will be endogenously determined by the researchers' choice of field. We also do not consider the possibility that gender preferences change in response to the gender ratio through a process where intergender coordination costs decline as individuals are more exposed to - and learn to communicate with - colleagues of the opposite sex. The first effect would tend to make women/men with a preferences for segregation select into fields with a higher/lower ratio of women to men and hence potentially make the female/male preference for segregation stronger in fields with more/less women. The effect of learning would run in the opposite direction. When splitting our sample into fields with high and low ratio of women, it appears if anything that the preference for segregation is higher for the articles in fields with a low ratio of women.

The paper proceeds as follows. In section 2, we develop a model of team formation where the prospective team members have preferences over team size and gender composition. We generate a set of testable predictions for how the male and female marginal propensities to form a team with a woman and to form a single team depend on the fraction of women in the pool of prospective team members. Next, we present data on team formation in the economics profession. In section 4, we test the predictions generated by the model using binomial Logit (and Probit) analysis of the probability to coauthor with a woman and the probability to single author. We are able to reject that the authorship pattern found in the data is consistent with gender neutral preferences. In fact, the data support an hypothesis stating that the fraction of the population (male and or female) of researchers preferring to join teams with colleagues of the same sex is larger than the fraction preferring gender mixed teams. Section 5 concludes.

2 Theoretical framework

Assume a population of male and female researchers within a research field. They can choose whether to do research alone or in teams. Assume that the proportion of female researchers is exogenously given by ϕ and the male proportion is $1 - \phi$. Researchers have preferences over teamwork versus working alone and over the gender of their team mates. It is assumed that preferences over team work and gender are independent, but we allow for the possibility that the distributions of preferences are gender specific.

Following Hamermesh and Oster (2002), the utility derived from teaming up in a particular constellation is presumably affected both by the expected productivity or quality of output of the team type, taking into account the returns to the specialization made possible by team work as well as the coordination costs that may arise in teams, and by the consumption value and costs of effort attached to the work process itself. We do not model explicitly how these possible productivity, consumption values and costs of effort result in preference rankings of team types and gender of team mates. Instead we assume that there is for each gender a distribution of agent types where the types are defined by how they rank team types in terms of team size and gender of team mates according to the utility they derive. We assume that an agent gets utility, u_A per project from a team type $A \in \{S_i, H, U_i\}$, where:

$$
\begin{aligned}
S_i &= \text{single team of sex } i \;, \\
H &= \text{heterogendered team, i.e. a team with the opposite sex,} \\
U_i &= \text{homogendered team of sex } i\text{, i.e. a team with the same sex.}
\end{aligned}
$$

Further, assume that agents rank a team according to the gender of the team mate. Some prefer working with their own sex (homo), some the opposite (hetero), while others are neutral. The reason for such ranking may be that there are gender specific real or perceived coordination cost that affect the quality of the output or the effort required in producing it. We allow the distribution of agents' gender rankings, or *gender preferences,* to be gender specific. Hence for each gender, i, there are three gender preference types (GP) distributed as follows:

Distribution of Gender Preferences (GP) of i		
fraction	Ranking of A	GP type
δ_i	$u_{U_i} > u_{H_i}$	Homo
γ_i	$u_{U_i} < u_{H_i}$	Hetero
$1 - \delta_i - \gamma_i$	$u_{U_i} = u_{H_i}$	Neutral

Table 1: Distribution of gender preferences.

Similarly, agents (or employees) rank team work differently. Again, team work can be more or less rewarding depending on the agent's ability to realize the potential returns to specialization or enjoy working in teams. We will assume that a fraction μ_i of gender i always ranks team work higher than working alone, while a fraction σ_i always prefers single work. The remaining $(1 - \sigma_i - \mu_i)$ fraction of agents will rank team work or single work higher depending on the sex of the prospective coauthor and depending on the gender preferences described above. Table 2 characterizes the distribution of team size preferences (TSP).

Distribution of Team Size Preferences (TSP) of gender i		
fraction	Preference ranking A	TSP type
σ_i	$u_{S_i} > \max\{u_{U_i}, u_{H_i}\}$	Single
μ_i	$u_{S_i} < \min\{u_{U_i}, u_{H_i}\}$	Multi
$1 - \sigma_i - \mu_i$	$\min\{u_{U_i}, u_{H_i}\} < u_{S_i} < \max\{u_{U_i}, u_{H_i}\}$	Conditional

Table 2: Distribution of team size preference.

It is assumed that $\{\sigma_i, \mu_i, \delta_i, \gamma_i\} \in [0,1]$, $\sigma_i + \mu_i \in [0,1]$, $\delta_i + \gamma_i \in [0,1]$ and, as mentioned above, that GP and TSP distributions are independent.

The joint distribution of gender and team size preferences for gender i, and the implied ranking of A is hence given by:

		Gender Preference		
		Homo	Neutral	Hetero
Team Size Preference		δ_i	$1-\delta_i-\gamma_i$	γ_i
Single	σ_i	$u_{S_i} > u_{U_i}$	$u_{S_i} > u_{U_i} = u_{H_i}$	$u_{S_i} > u_{H_i}$
Conditional	$1-\sigma_i-\mu_i$	$u_{H_i} < u_{S_i} < u_{U_i}$	$u_{H_i} = u_{S_i} = u_{U_i}$	$u_{U_i} < u_{S_i} < u_{H_i}$
Multi	μ_i	$u_{S_i} < u_{H_i}$	$u_{S_i} < u_{H_i} = u_{U_i}$	$u_{S_i} < u_{U_i}$

Table 3: Joint distribution of gender and team size preferences.

2.1 Team formation and the distribution of projects by team type

At each period in time, agents are randomly grouped in pairs. Each pair will form a team and jointly work on two projects (articles) unless one or both strictly prefer working alone. In that case, they will both constitute a single team each. Hence, the resulting output in number of projects per person is constant across team types.[3] Given the distribution of preferences described above, we can derive the distribution of projects by type of team. This matching process is, of course, totally mechanical in the sense that individual researchers do not act on their preference in order to affect the likelihood of finding a good match. One way to look at it, is that by choosing to work within a specific field, where the distribution of preferences and gender composition are those described above, the individual has already maximized the likelihood of finding the preferred match. The matching process described here then takes place within the agents' preferred field and results in a distribution of performed projects or articles by team types.

The fraction of homogendered female teams is then determined by the probability of two women are matched in a pair and the probability that none of the two strictly prefers working alone to working with another woman, that is:

[3] In the economics literature, this assumption is supported empirically in the findings of Sauer (1988) and McDowell and Kiholm Smith (1992).

$$P(U_f) = \phi^2 \left(\mu_f + (1-\gamma_f)(1-\sigma_f-\mu_f)\right)^2.$$

The same reasoning applies throughout. The share of homogendered male teams:

$$P(U_m) = (1-\phi)^2 \left(\mu_m + (1-\gamma_m)(1-\sigma_m-\mu_m)\right)^2.$$

The share of heterogendered teams:

$$P(H) = 2\phi\,(1-\phi)\left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right).$$

The share of female single teams:

$$\begin{aligned} P(S_f) &= \phi^2\left(1-\left(\mu_f + (1-\gamma_f)(1-\sigma_f-\mu_f)\right)^2\right) + \\ &\quad \phi\,(1-\phi)\left(1-\left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right)\right) \end{aligned}$$

The share of male single teams:

$$\begin{aligned} P(S_m) &= (1-\phi)^2\left(1-\left(\mu_m + (1-\gamma_m)(1-\sigma_m-\mu_m)\right)^2\right) + \\ &\quad \phi\,(1-\phi)\left(1-\left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right)\right). \end{aligned}$$

It is easily verified that the total share of female teams, $P(F)$ is indeed ϕ :

$$P(F) = P(U_f) + P(S_f) + \frac{1}{2}P(H) = \phi,$$

and that the sum of all types adds up to unity.

2.2 Implications for differences in male and female team patterns

We can now derive implications for how the propensity to work alone will differ for male and female agents as well as implications for male-female differences in the marginal propensity to team up with a woman. The model generates different implications depending on our assumptions regarding the distributions of preferences. The aim is to generate testable predictions for how the population's gender composition ought to affect the male and female propensity to form a single team and to team up with a woman.

The marginal propensities to have a female team mate (FTM), given that one is a man or woman, are:

$$P(FTM \mid m) = \frac{P(H)}{2\,(1-\phi)} = \beta_{FTM_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m)\ \phi$$
$$P(FTM \mid f) = \frac{P(U_f)}{\phi} = \beta_{FTM_f}(\sigma_f, \mu_f, \delta_f, \gamma_f)\ \phi,$$

where the coefficients β_{FTM_m} and β_{FTM_m} are given by:

$$\begin{aligned} & \beta_{FTM_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m) \\ = & \left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right), \\ \beta_{FTM_f}(\sigma_f, \mu_f, \gamma_f) = & \left(\mu_f + (1-\gamma_f)(1-\sigma_f-\mu_f)\right)^2. \end{aligned}$$

Marginal propensities to single team given that one is male or female are:

$$P(S \mid m) = \frac{P(S_m)}{(1-\phi)} = \alpha_m\left(\sigma_m, \mu_m, \gamma_m\right) + \beta_{S_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m, \gamma_m)\phi$$
$$P(S \mid f) = \frac{P(S_f)}{(\phi)} = \alpha_f(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m) + \beta_{S_f}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m, \gamma_f)\phi$$

where the coefficient, α_i and β_{S_i} are given by:

$$\alpha_m\left(\sigma_m, \mu_m, \gamma_m\right) = \left(1 - (\mu_m + (1-\gamma_m)(1-\sigma_m-\mu_m))^2\right),$$

$$\begin{aligned} & \beta_{S_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m, \gamma_m) \\ = & (\mu_m + (1-\gamma_m)(1-\sigma_m-\mu_m))^2 - \\ & \left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right) \end{aligned}$$

$$\begin{aligned} & \alpha_f(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m) \\ = & \left(1 - \left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right)\right) \end{aligned}$$

$$\begin{aligned} & \beta_{S_f}(\sigma_f, \sigma_m, \mu_f, \mu_m, \delta_f, \delta_m, \gamma_f) \\ = & \left(\mu_f + (1-\delta_f)(1-\sigma_f-\mu_f)\right)\left(\mu_m + (1-\delta_m)(1-\sigma_m-\mu_m)\right) \\ & - \left(\mu_f + (1-\gamma_f)(1-\sigma_f-\mu_f)\right)^2. \end{aligned}$$

We can now turn to the implications for β_{FTM_i}, α_i and β_{S_i} of alternative hypotheses regarding the distributions of preferences of male and female employees.

We will formulate hypotheses that depart little by little from the idea that gender is irrelevant, i.e. that there is only one population of gender neutral agents who are all drawn from the same distribution of team size preferences. This, will be our first hypothesis:

Hypothesis 1: *Gender irrelevance*, i.e. there is only one population. i.e. $\sigma_f = \sigma_m = \sigma$, $\mu_f = \mu_m = \mu$, and $\delta_f = \delta_m = \gamma_f = \gamma_m = 0$.

$$\begin{aligned} \beta_{FTM_m}(\sigma, \sigma, \mu, \mu, 0, 0) &= (1-\sigma)^2 \\ \beta_{FTM_f}(\sigma, \mu, 0) &= (1-\sigma)^2 \end{aligned}$$

$$\begin{aligned} \alpha_m(\sigma, \mu, 0) &= 2\sigma - \sigma^2 \\ \beta_{S_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, 0, 0, 0) &= 0 \end{aligned}$$

$$\begin{aligned} \alpha_f(\sigma, \sigma, \mu, \mu, 0, 0) &= 2\sigma - \sigma^2 \\ \beta_{S_f}(\sigma_f, \sigma_m, \mu_f, \mu_m, 0, 0, 0) &= 0 \end{aligned}$$

The implication of gender irrelevance is that there should be no differences in the male and female marginal propensities to team up with a woman or to work alone, that is:

$$\begin{aligned} \beta_{FTM_m} &= \beta_{FTM_f} \\ \alpha_m &= \alpha_f \\ \beta_{S_m} &= \beta_{S_f} = 0. \end{aligned}$$

Our second hypothesis, assumes gender neutrality, but allows men and women to be drawn from different team size preference distributions:

Hypothesis 2: *Gender neutrality,* i.e. the fraction of men and women who prefer single work and team work are different, but all employees are gender neutral: $\delta_f = \delta_m = \gamma_f = \gamma_m = 0$.

$$\begin{aligned} \beta_{FTM_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, 0, 0) &= (1-\sigma_f)(1-\sigma_m) \\ \beta_{FTM_f}(\sigma_f, \mu_f, 0) &= (1-\sigma_f)^2 \end{aligned}$$

$$\begin{aligned} \alpha_m(\sigma_m, \mu_m, 0) &= 1-(1-\sigma_m)^2 \\ \beta_{S_m}(\sigma_f, \sigma_m, \mu_f, \mu_m, 0, 0, 0) &= (1-\sigma_m)(\sigma_f - \sigma_m) \end{aligned}$$

$$\begin{aligned}\alpha_f(\sigma_f,\sigma_m,\mu_f,\mu_m,0,0) &= \sigma_f+\sigma_m-\sigma_f\sigma_m\\ \beta_{S_f}(\sigma_f,\sigma_m,\mu_f,\mu_m,0,0,0) &= (1-\sigma_f)(\sigma_f-\sigma_m).\end{aligned}$$

When there is gender neutrality, the relative size of the fraction of women and men who prefer single work will determine the relative sizes of the coefficients.

$$\begin{aligned}\beta_{FTM_m} &\lesseqgtr \beta_{FTM_f} \Leftrightarrow \sigma_m \gtreqless \sigma_f\\ \alpha_m &\gtreqless \alpha_f \Leftrightarrow \sigma_m \gtreqless \sigma_f\\ \beta_{S_m},\beta_{S_f} &\lesseqgtr 0 \Leftrightarrow \sigma_m \gtreqless \sigma_f,\text{ and } \beta_{S_m} \gtreqless \beta_{S_f} \Leftrightarrow \sigma_m \gtreqless \sigma_f.\end{aligned}$$

Our third hypothesis maintains that men and women have the same team size preferences, but imposes that they also have non neutral and the same preferences for gender. That is, the fraction of men how prefer to team up with men (women) is the same as the fraction of women who prefer to team up with men (women):

Hypothesis 3: *Identical but not neutral*, i.e. men and women have the same team preferences and gender preferences: $\sigma_f=\sigma_m=\sigma$, $\mu_f=\mu_m=\mu$, $\delta_f=\gamma_m=f$, $\delta_m=\gamma_f=m$.

$$\begin{aligned}\beta_{FTM_m}(\sigma,\sigma,\mu,\mu,f,m) &= (1-\sigma-f(1-\mu-\sigma))\,(1-\sigma-m(1-\mu-\sigma))\\ \beta_{FTM_f}(\sigma,\mu,m) &= (1-\sigma-m(1-\mu-\sigma))^2\end{aligned}$$

$$\begin{aligned}\alpha_m\,(\sigma,\mu,f) &= 1-(1-\sigma-f(1-\mu-\sigma))^2\\ \beta_{S_m}(\sigma,\sigma,\mu,\mu,f,m,f) &= (1-\sigma-f(1-\mu-\sigma))^2\\ &\quad-(1-\sigma-f(1-\mu-\sigma))\,(1-\sigma-m(1-\mu-\sigma))\end{aligned}$$

$$\begin{aligned}\alpha_f(\sigma,\sigma,\mu,\mu,f,m) &= 1-(1-\sigma-f(1-\mu-\sigma))\,(1-\sigma-m(1-\mu-\sigma))\\ \beta_{S_f}(\sigma,\sigma,\mu,\mu,f,m,m) &= (1-\sigma-f(1-\mu-\sigma))\,(1-\sigma-m(1-\mu-\sigma))\\ &\quad-(1-\sigma-m(1-\mu-\sigma))^2\end{aligned}$$

The implications of identical but not neutral preferences is, as with hypothesis 2, that only some combinations of the relative sizes of the coefficients are consistent:

$$\begin{aligned}\beta_{FTM_m} &\lesseqgtr \beta_{FTM_f} \Leftrightarrow f \gtreqless m\\ \alpha_m &\lesseqgtr \alpha_f \Leftrightarrow m \gtreqless f\\ \beta_{S_f} &< \beta_{S_m}.\end{aligned}$$

So far we have considered hypotheses that do not allow for a preference for segregation. The next three hypotheses all imply that there is a possibility both for preference for segregation and integration.

Our fourth hypothesis assumes that men and women have symmetrical gender preferences, but the same preferences for team size. The share of women who prefer men (women) is the same as the share of men who prefer women (men). We will define an over all preference for segregation as a situation where the fraction preferring the own sex is larger than the fraction preferring to work with the opposite sex.

Hypothesis 4: *Mutual gender preference*, i.e. men and women have the same team preferences and symmetrical gender preferences: $\sigma_f = \sigma_m = \sigma$, $\mu_f = \mu_m = \mu$, $\delta_f = \delta_m = \delta$, $\gamma_f = \gamma_m = \gamma$.

$$\begin{aligned}\beta_{FTM_m}(\sigma,\sigma,\mu,\mu,\delta,\delta) &= (1-\sigma-\delta(1-\mu-\sigma))^2 \\ \beta_{FTM_f}(\sigma,\mu,\gamma) &= (1-\sigma-\gamma(1-\mu-\sigma))^2\end{aligned}$$

$$\begin{aligned}\alpha_m(\sigma,\mu,\gamma) &= 1-(1-\sigma-\gamma(1-\mu-\sigma))^2 \\ \beta_{S_m}(\sigma,\sigma,\mu,\mu,\delta,\delta,\gamma) &= (1-\sigma-\gamma(1-\mu-\sigma))^2-(1-\sigma-\delta(1-\mu-\sigma))^2\end{aligned}$$

$$\begin{aligned}\alpha_f(\sigma,\sigma,\mu,\mu,\delta,\delta) &= 1-(1-\sigma-\delta(1-\mu-\sigma))^2 \\ \beta_{S_f}(\sigma,\sigma,\mu,\mu,\delta,\delta,\gamma) &= (1-\sigma-\delta(1-\mu-\sigma))^2-(1-\sigma-\gamma(1-\mu-\sigma))^2\end{aligned}$$

Mutual gender preferences put restrictions on all the relative sizes of coefficients. Moreover, the direction of the inequalities will be informative as to whether we have an overall preference for gender segregation or not.

$$\begin{aligned}\beta_{FTM_m} &\lesseqqgtr \beta_{FTM_f} \Leftrightarrow \delta \gtreqqless \gamma \\ \alpha_m &\lesseqqgtr \alpha_f \Leftrightarrow \delta \gtreqqless \gamma \\ \beta_{S_f} &\lesseqqgtr \beta_{S_m} \Leftrightarrow \delta \gtreqqless \gamma\end{aligned}$$

Hypotheses 5 and 6 are versions of the same hypothesis, namely that one gender is neutral while the other has gender preferences as described in hypothesis 4.

Hypothesis 5: *Female gender preferences*, i.e. men and women have the same team preferences, men are gender neutral, women are not: $\sigma_f = \sigma_m = \sigma$, $\mu_f = \mu_m =$

μ, $\delta_f \neq \delta_m = 0$, $\gamma_f \neq \gamma_m = 0$.

$$\begin{aligned}
\beta_{FTM_m}(\sigma,\sigma,\mu,\mu,\delta_f,0) &= (1-\sigma)\,(1-\sigma-\delta_f(1-\mu-\sigma)) \\
\beta_{FTM_f}(\sigma,\mu,\gamma_f) &= \big(1-\sigma-\gamma_f(1-\mu-\sigma)\big)^2
\end{aligned}$$

$$\begin{aligned}
\alpha_m\left(\sigma,\mu,\gamma_f\right) &= 1-(1-\sigma)^2 \\
\beta_{S_m}(\sigma,\sigma,\mu,\mu,\delta_f,0,\gamma_f) &= (1-\sigma)^2-(1-\sigma)\,(1-\sigma-\delta_f(1-\mu-\sigma))^2
\end{aligned}$$

$$\begin{aligned}
\alpha_f(\sigma,\sigma,\mu,\mu,\delta_f,0) &= 1-(1-\sigma)\,(1-\sigma-\delta_f(1-\mu-\sigma)) \\
\beta_{S_f}(\sigma,\sigma,\mu,\mu,\delta_f,0,\gamma_f) &= (1-\sigma)\,(1-\sigma-\delta_f(1-\mu-\sigma))^2-\big(1-\sigma-\gamma_f(1-\mu-\sigma)\big)^2
\end{aligned}$$

Female gender preferences restricts $\alpha_m \leq \alpha_f$. The directions of the other inequalities are then informative of whether there is a preference for segregation.

$$\begin{aligned}
\beta_{FTM_m} &\lesseqgtr \beta_{FTM_f} \Leftrightarrow \delta_f \gtreqless 2\gamma_f-\gamma_f^2\frac{(1-\mu-\sigma)}{(1-\sigma)} \\
\alpha_m &\leq \alpha_f \Leftrightarrow \delta_f \geq 0 \\
\beta_{S_f} &\lesseqgtr \beta_{S_m} \Leftrightarrow \delta_f \gtreqless \gamma_f-\gamma_f^2\frac{(1-\mu-\sigma)}{2(1-\sigma)}
\end{aligned}$$

Hypothesis 6: *Male gender preferences*, i.e. men and women have the same team preferences, women are gender neutral, men are not: $\sigma_f = \sigma_m = \sigma$, $\mu_f = \mu_m = \mu$, $0 = \delta_f \neq \delta_m$, $0 = \gamma_f \neq \gamma_m$.

$$\begin{aligned}
\beta_{FTM_m}(\sigma,\sigma,\mu,\mu,0,\delta_m) &= (1-\sigma)\,(1-\sigma-\delta_m(1-\mu-\sigma)) \\
\beta_{FTM_f}(\sigma,\mu,0) &= (1-\sigma)^2
\end{aligned}$$

$$\begin{aligned}
\alpha_m\left(\sigma,\mu,\gamma_m\right) &= 1-(1-\sigma-\gamma_m(1-\mu-\sigma))^2 \\
\beta_{S_m}(\sigma,\sigma,\mu,\mu,0,\delta_m,\gamma_m) &= (1-\mu-\sigma)\,\big((1-\sigma)(\delta_m-2\gamma_m)+(1-\mu-\sigma)\gamma_m^2\big)
\end{aligned}$$

$$\begin{aligned}
\alpha_f(\sigma,\sigma,\mu,\mu,0,\delta_m) &= (2-\sigma)\sigma+\delta_m(1-\sigma)(1-\mu-\sigma) \\
\beta_{S_f}(\sigma,\sigma,\mu,\mu,\delta_f,0,\gamma_f) &= \delta_m(1-\sigma)(1-\mu-\sigma)
\end{aligned}$$

Male gender preferences implies that $\beta_{FTM_m} \leq \beta_{FTM_f}$. The other inequalities are then informative of if there is a preference for gender segregation.

$$\begin{aligned}
\beta_{FTM_m} &\leq \beta_{FTM_f} \Leftrightarrow \delta_m \geq 0 \\
\alpha_m &\lesseqgtr \alpha_f \Leftrightarrow \delta_m \gtreqless 2\gamma_m-\gamma_m^2\frac{(1-\mu-\sigma)}{(1-\sigma)} \\
\beta_{S_f} &\lesseqgtr \beta_{S_m} \Leftrightarrow \delta_m \gtreqless \gamma_m-\gamma_m^2\frac{(1-\mu-\sigma)}{2(1-\sigma)}
\end{aligned}$$

It is clear from the model, that each hypothesis regarding preferences places a set of restrictions on the parameters of the model. We now proceed to investigate if any of the proposed hypotheses can be rejected by data on coauthorship patterns within the field of economics.

3 Team formation in Economics

Team formation in academia and in economics is voluntary. Researchers team up when both parties to the team think they are better off doing that than writing alone. A deviation from gender neutrality in the pattern of coauthorship in articles published in academic journals, would hence reveal that there is something about gender which affects researcher willingness to coauthor.

We focus on three general top journals, American Economic Review, Journal of Political Economy and Quarterly Journal of Economics. This is a very restrictive sample of journals that has been chosen to control for the quality of the teams formed.[4]

Arguably the publication patterns in these top journals reflect the coauthorship behavior of economists at the very top of the ability distribution. Unless it is possible to make a case that the male and female ability distributions and/or ambitions to publish at the top are markedly different, we will argue that our results will reveal the team and gender preferences of the top tier of economists. Even if men and women were to have different cut-off quality levels for when they judge a paper good enough to send to these journals, the very low success rate is likely to undo any such biases since, the cut-off quality for publication is likely to exceed the cut-off quality for submission for all types of authorships. Another concern is that the refereeing process does discriminate against some types of authorships. According to Blank (1991) there is no evidence of bias against any sex in the referee process of these journals.

Economics and many other academic fields have seen an increase in the prevalence of teamwork. Hamermesh and Oster (2002) report an average share of coauthored articles of 30 per cent in 1970-1979 period in the American Economic Review, the Journal of Political Economy and the Quarterly Journal of Economics. For the

[4] A further reason to select these three journals is that the authorship patterns in them have previously been studied in Laband and Tollison (2000) and Hamermesh and Oster (2002).

1990-2002 period, we find that this share has increased to 55 per cent.[5]

The field of economics is very male dominated. According to the Annual Reports of the American Economic Association Committee on the Status of Women in the Economics Profession, the female share of faculty at Ph.D. granting departments was 10 per cent in 2000, a doubling since 1976.[6] Furthermore, Ferber and Teiman (1980) and McDowell and Kiholm Smith (1992) establish that there is a pattern of gender sorting in coauthorships in the field of economics and that women write more single authored papers. In data on the publications during the 1969-1986 period of a sample of 178 Ph.D.'s from Top20 US institutions, McDowell and Kiholm Smith (1992) report a significant gender difference in the propensity to coauthor with a woman. In particular, they found that a woman was more than 5 times as likely as man to coauthor with a woman. McDowell and Kiholm Smith (1992) also propose that gender sorting in coauthorships may be one reason for why women have had a hard time getting ahead in economics, arguing that the small fraction of women in economics makes the pool of potential coauthors limited for women.[7]

Clearly, if coauthorship has become more productive over time, and women have a hard time finding coauthors due to gender sorting, the returns to pursuing a career in economics are low for women and increasingly so. However, the gender sorting observed in previous studies need not reflect that men and women prefer gender segregated teams. Is is possible that the previously observed gender sorting is a result of gender differences in fields of interest, since it is well known and confirmed in our data, that the share of women varies across subfields of economics. It is also possible that men's and women's perceptions of the returns and costs to forming teams may vary across fields.

[5] Hamermesh and Oster (2002) find evidence that the increase prevalence of teamwork is indeed associated with an increase in the relative productivity of coauthored as opposed to single authored papers as measured by citations. It seems reasonable to conclude that increased returns to specialization has been part of the driving force.

[6] Including untenured faculty in the figure improves the female share in 2000 to 15 per cent. However, restricting the sample to the Top20 deparments puts the female share at 8 per cent (tenured faculty) and 13 (untenured included).

[7] There are a number of papers discussing the difficulties for women in the economics profession, e.g. Kahn (1995, 2002), McDowell, Singell and Ziliac (1999, 2001).

3.1 The Data

The data, downloaded from EconLit, stretches from 1991 to 2002 and comprises all issues of these three journals, in total 3140 articles.[8] The American Economic Review has by far the most entries, 1965 compared to 601 for the JPE, and 509 for the QJE. Other than that, there are no significant differences between the journals, except for the American Economic Review on average having shorter articles than the other two journals.[9] (Summary statistics for each journal are reported in Appendix A.)

We have gender coded the authors. Excluding the 65 (2.07 per cent) articles where we have been unable to identify the sex of at least one of the contributing authors, we remain with a data set consisting of 3075 articles. Throughout our analysis, each article accounts for one *authorship*. For example, let there be two female-authored articles in an issue, one is single-authored and the other is coauthored with a male colleague. While the number of female authors in that issue obviously is 2, the number of female authorships is 1.5, that is 1 single authorship plus 0.5 of a doubled authored article.

The authorship pattern by gender and team size is described in Table 1, where it is clear that although the total female share of articles is 13 per cent, for the time period as a whole the differences in the distributions of authorship types by gender are small. Thus, it is no longer obvious that women have a harder time finding coauthors than men, as was the case in the 1980s according to McDowell and Kiholm Smith (1992).The JEL-codes provided by EconLit are used to classify the articles into fields.

The average number of JEL-codes per article is 1.44. Hence each article can belong to more than one field. The nine most frequent JEL-codes each accounts for more than five per cent of the articles every year of our sample period.[10]

[8] For 2002, we do not have the last numbers of the JPE and QJY in the sample and we also lack the last two numbers of the AER since they were not available in Econlit when we collected our data. 1991 is the first year in our sample due to the reform of the JEL classification that occurred that year. To avoid mis-classifications we have not attempted to recode the JEL-classifications for articles published in 1990 and earlier.

[9] The AER has more characters per page than the other two journals, so that the difference is somewhat smaller than it seems.

[10] In no single year does any of the other JEL-codes rank higher than these top nine.

Type of authorship	Gender		
	female	male	all
single	178	1167	1345
row %	*13*	*87*	
col %	45	44	44
double	177	1155	1332
row %	*13*	*87*	
col %	45	43	43
multiple	40	358	398
row %	*10*	*90*	
col %	10	13	13
all	395	2680	3075
	13	*87*	*100*

Table 4: Distribution of authorship types by gender.

These JEL-codes define the fields of economics that will be analyzed throughout the paper, Micro Economics(D), Macroeconomics(E), International Economics(F), Financial Economics(G), Public Economics(H), Health and Education(I), Labor Economics(J), Industrial Organization(L), and Growth and Development(O). The remaining JEL-codes are lumped together in an Other-category (ZZ). The relative frequency of distribution of JEL-codes is shown as stacked bars in Figure 1, which also displays the share of female authorships in the major fields' respective bar.

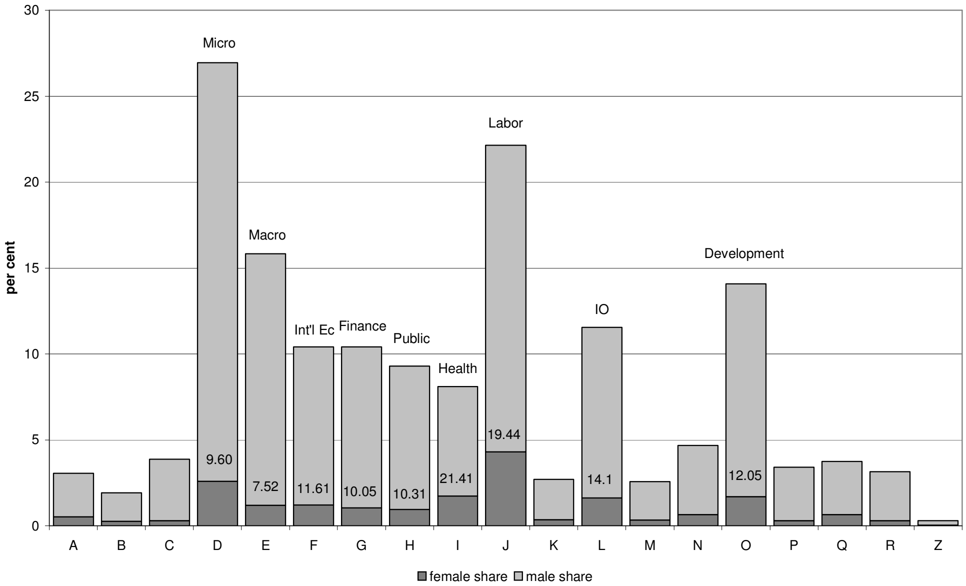

Figure 1: The relative frequency of JEL-codes and the female share of articles.

Clearly, there are large differences the share of female articles across field. Female authorships are roughly three times more frequent in Health and Education (21.4%)

than in Macroeconomics (7.5%). Figure 2 illustrates the absolute presence in number of articles of the major field and in each bar the figure shows that there are field differences also in the relative frequency of single and coauthorship. We will think of these differences as reflecting possible differences in technology across fields resulting in differences in the returns and costs of specialization.

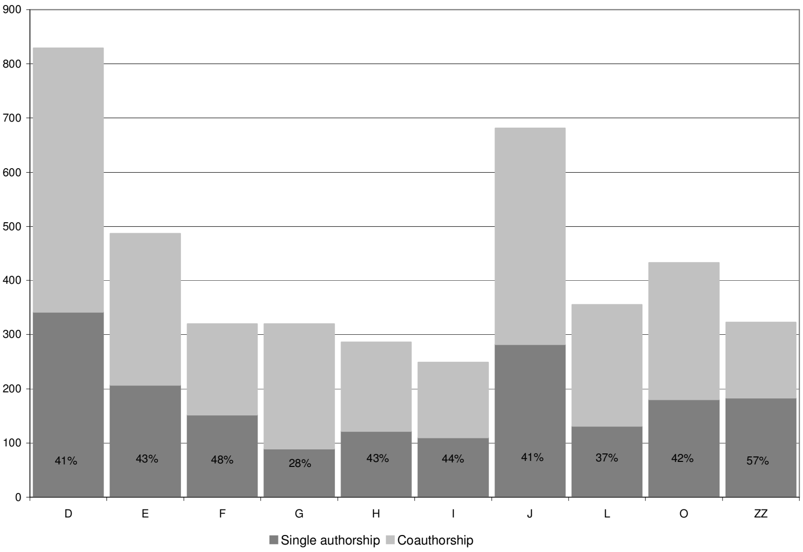

Figure 2: The size of fields and the share of single and coauthored articles.

3.2 The propensity to single author and gender sorting in coauthorship

Before we go on with formal testing of the hypotheses generated by the model, it is warranted to take a glance at what the data tell us about the relation between the fraction of women within a field and gender differences in single authorship on the one hand and the propensity to coauthor with a woman on the other. First, Figure 3 shows how the female-male gap in the propensity to single author relates to the share of female authorships in a subfield of economics. It becomes clear that at the field level, only in health and education and in public economics women single author less than men. It is possible to draw a negative trend line, i.e. that the gap shrinks with the share of women, but the relationship is not at all a clear one and it could in fact be that there are two groups, each with a positive slope.

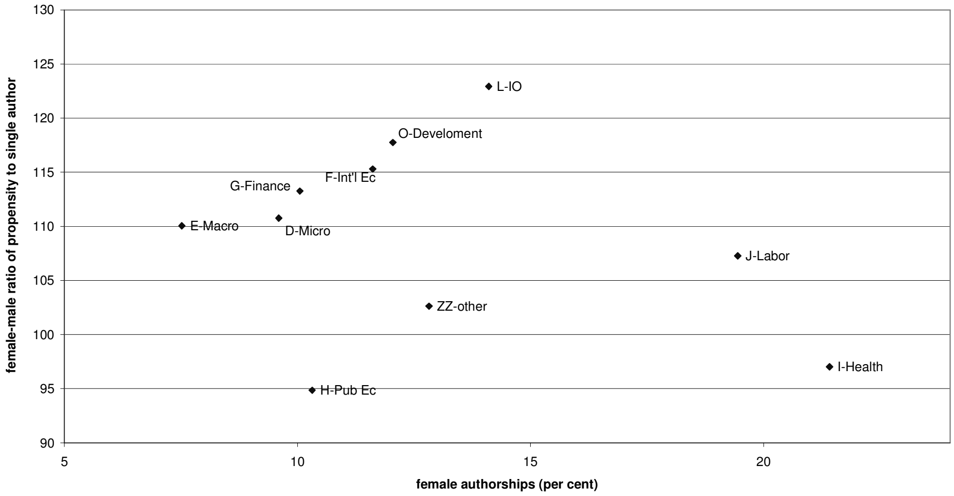

Figure 3: The female-male gap in the propensity to single author and the share of female authorships, by JEL-code.

Looking instead at the propensity of having a female coauthor, Figure 4 reveals that the higher the share of women in a certain field, the larger is the difference between female and male coauthorships patterns. As the share of female authorships increases in a field, women tend to increasingly write with other women, while men's coauthoring patterns, although affected, are much less.

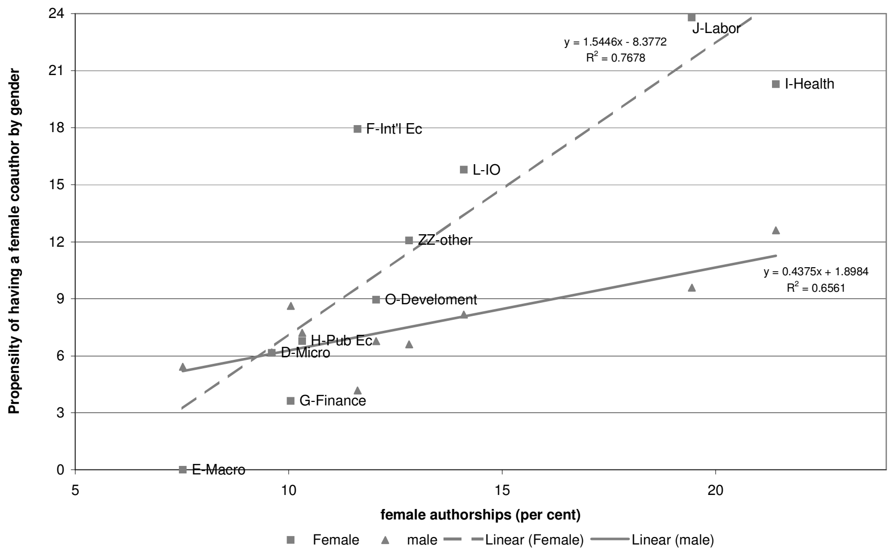

Figure 4: Female authorships and propensity of having a female coauthor by gender, by JEL-codes.

This first look at the data does not, at first hand, lend support to the hypothesis of gender neutral team formation. However, these scatter plots do not control for the fact that women and men in our data are of different seniority, that there have been changes in the returns to coauthorship over time, etc. In fact, according to our model the pattern in Figure 4 would be consistent with gender neutrality if it was the case that the gap in single authorship is reversed should we control for that men are more senior. Hence, we cannot rule out that team formation in economics is gender neutral, yet.

4 Testing gender neutrality

The aim of this section is to formally test whether there are any revealed gender preferences in the pattern of coauthorship in articles published in the American Economic Review, the Journal of Political Economy and the Quarterly Journal of Economics between 1991 and 2002. We do this by studying differences in the propensity of men and women having a female coauthor, and differences in the propensity to single author, and how these depend on the share of women within academic fields using binomial logit and probit analysis. The model presented in section 2 above, provides clear predictions for how gender differences in authorship patterns ought to depend on the gender ratio within a field. If we are able to reject hypotheses 1,2,3, and if we cannot reject 4,5 or 6, we will argue that we have indeed found evidence in support of a preference for gender segregation.

The 3075 articles are the observations of authorships. We view each authorship from the point of view of the first author of the article. Doing so we use the alphabet and family names to draw a representative author for each article. It is the behavior of this representative author that we study in order to establish how the fraction of women in a field influences the authorship behavior of male and female authors when we control for seniority effects, time trends, for possible technology differences across fields of economic and some article specific measures such as journal dummies, number of pages, belonging to the May issue of the AER, and so on.

4.1 Econometric specification

The model developed in section 2 resulted in the following simple equations:

$$\begin{aligned} P(FTM_{is}) &= \beta_{iFTM_s}\,\phi_i \\ P(S_{is}) &= \alpha_{is} + \beta_{iS_s}\phi_i\ . \end{aligned}$$

That is, the probability of having a female coauthor and the probability of single authoring in field i for an author of gender s are linear functions of the fraction of women in the field, where the coefficients depend of the preference distributions of the male and female researchers within the field, where we need to allow for the fact that the distributions of preferences depend on field technology that may vary over time and on author characteristics (other than gender) such as seniority. Hypotheses 1-6 placed a number of testable restrictions on the relative sizes of the male and female $\beta_{iFTM_s}, \beta_{iS_s}$ and α_{is} coefficients.

We take our model to the data by estimating the following two equations. The first equation which is estimated is the probability that a researcher j in field i coauthors with a woman:

$$\begin{aligned} &P(FTM)_{ji} = \beta_{FTM1} + \beta_{FTM2}sex_j +\ \beta_{FTM3}\phi_i + \beta_{FTM4}seniority_j + \\ &+\beta_{FTM5}(print\ year) + sex_j\ (\beta_{FTM6}\phi_i + \beta_{FTM7}seniority_j + \beta_{FTM8}(print\ year)), \end{aligned}$$

where sex_j is a dummy for gender. The key coefficients of interest are β_{FTM3} and β_{FTM6} which capture how the fraction of women ϕ_i in a field affects the marginal propensity, β_{iFTM_s} of men and women to coauthor with a woman. The second equation is the probability to single author of researcher j in field i:

$$\begin{aligned} &P(S)_{ji} = \alpha_{S1} + \alpha_{S2}sex_j + \\ &\ \beta_{S3}\phi_i + \beta_{S4}seniority_j + \beta_{S5}(print\ year) +\ \beta_{S6}\,(returns\ to\ specialization)_i + \\ &sex_j\ (\beta_{S7}\phi_i + \beta_{S8}seniority_j + \beta_{S9}(print\ year) + \beta_{S10}\,(returns\ to\ specialization)_i)\quad . \end{aligned}$$

The key coefficients of interest in this equation are: α_{S1} and α_{S2}, which capture gender differences in levels of single authorship (α_{is}); β_{S3} and β_{S7} which capture gender differences in the sensitivity of single authorship to the fraction of women (β_{iS_s}). In what follows, we discuss the construction of variables and some additional control variables.

4.2 Variable construction

The dependent variable of or first equation is the *Probability of having a female coauthor*. The dependent variable takes the value one if at least one of the second through eighth authors of the article is a woman, otherwise the dependent variable takes the value zero. That is, if the first author is female, the article needs to have at least two female authors altogether. This implies, of course, that the dependent variable takes the value zero for all single authored articles. Most coauthored articles have only two authors. There are however a number of papers with more than two authors and for these the true probability of having a female coauthor is not binary. To account for this, we introduce a dummy variable for articles with *more than two authors*. In the second set of regressions, the dependent variable is single authorship, i.e. the dependent variable takes the value one if the article has one author and zero otherwise.

As we take the model to the data, we go from an analysis within a specific field of economics to an analysis across fields and we need to handle that each article can have more than one field-code. To capture how coauthorship patterns vary with the female presence in the fields of the article, we construct an article specific measure of the *mean female share*, i.e. a measure of the fraction of women in the population of researchers, ϕ_i, relevant to the article. This measure is calculated as the average of the female share of authorships of the fields corresponding to the JEL-codes of the article.[11] In order to control for possible differences in technology, i.e. in returns and costs of teamwork, we construct a analogous article specific measure for the *mean share of single authored articles.* Throughout, we also enter dummy variable for each JEL-field to control for other field specific differences.

Since our data is on articles and not on authors, we have very little author specific information. The purpose of the analysis is of course to control for gender, which we do by controlling for the *sex* of the first author, which takes the value one if it is a woman and zero otherwise. Recognizing that there may be gender differences in the seniority distributions of our representative first authors we construct two measures of seniority. Clearly, a reason why women do not coauthor to the same extent as

[11] The *female share of authorships* of a field X is the average female share of authorships over the whole sample period for all articles with the JEL-code X. Hence, if, for example, an article has two JEL codes in E-Macro, such as E120 and E401, and one in J-Labour, the *mean female share* of the article is the average of the *female share of authorships* in macro and labor.

men can be that they are less well established. The first measure of seniority is simply the *number of publications* in our data of the individual author up to the year of the publication of the article. A second measure related to seniority, tries to capture seniority gaps within the authorship team. More specifically, the variable *seniority gap* is defined as the gap between the number of publications of the first author and the coauthor with least publications.

Female and male authors might also be affiliated to more or less top universities. Since authors' affiliation according to Kahn (1995) is the most important determinant of differences in men's and women's publication records, we control for gender differences in affiliation. Therefore, we have classified the affiliations of the authors in our sample according to the share of publications they account for. Three universities, University of Chicago, Harvard University and MIT, each accounts for more than 5 per cent of total publications during the entire period 1991-2002. Six other universities have authors that make up more than 2 per cent of the publications (Berkeley, Columbia, Michigan, Northwestern, Princeton and Stanford). The dummy *Top 9 university* assumes the value of one if the first author is affiliated with one of the nine universities that have more then 2 per cent of all publications, and 0 otherwise.

We will also control for the *print year* as a means to control for time trends. Controls for *source (i.e. the Journal)* of the article, the *number of pages* are also used.

4.3 Results

We start by running binomial logits to estimate the probability of having a female coauthor and of single-authoring, given that the first author is a man or a woman. The results are presented in Table 5. (The corresponding probit estimates are reported in Appendix B.)

	(1)	(2)	(3)	(4)
	Probability of having a female coauthor		Probability of single-authoring	
Sex	0.879***	19.372	-0.016	181.792**
	(0.171)	(104.618)	(0.115)	(71.049)
Mean female share	10.618**	6.854	-4.183	-3.363
	(4.844)	(4.902)	(2.609)	(2.696)
Sex * (Mean female share)		12.211***		-5.118*
		(4.216)		(2.949)
Print year	0.057***	0.059***	-0.040***	-0.028**
	(0.022)	(0.025)	(0.012)	(0.013)
Sex * (Print Year)		-0.010		-0.091**
		(0.052)		(0.036)
Number of publications	0.052	0.049	-0.103***	-0.099***
	(0.032)	(0.034)	(0.020)	(0.020)
Sex * (N. of publications)		0.055		-0.175*
		(0.096)		(0.095)
Top 9 university	-0.267	-0.296	0.529**	0.505***
	(0.181)	(0.202)	(0.087)	(0.092)
Sex * (Top 9 university)		0.107		0.374
		(0.437)		(0.289)
Mean single			2.231***	1.986**
			(0.837)	(0.849)
More than two authors	1.431***	1.445***		
	(0.165)	(0.166)		
Number of pages	-0.014*	-0.014*	-0.023***	-0.023***
	(0.008)	(0.008)	(0.005)	(0.005)
AER	0.007	0.004	0.193	0.194
	(0.213)	(0.215)	(0.121)	(0.121)
JPE	-0.007	-0.027	0.028	0.037
	(0.262)	(0.265)	(0.139)	(0.140)
Constant	-117.72***	-122.10***	79.829***	56.395**
	(44.062)	(48.963)	(23.395)	(24.919)
Observations	3072	3072	3072	3072
Wald Chi^2	148.33	165.91	174.28	187.81
Prob> Chi^2	0.000	0.000	0.000	0.000
Pseudo R^2	0.0916	0.0972	0.0462	0.0502

Robust standard errors in parentheses. * denotes significance at least at the 90 per cent level; ** significance at least at the 95 per cent level; and *** significance at least at the 99 per cent level. All estimations include dummy variables for JEL-codes (not reported).

Table 5: Basic binomial logit results.

Columns (1) and (2) report the logit estimations of the probability of having a female coauthor with and without interaction effects. If no interaction terms are included, women are significantly more likely to have a female coauthor than men, and all authors are more likely to write with a woman as the mean female share of authorships increases in the subfields of the article. The probability of having a female coauthor also rises significantly over time, while not being affected by the author's number of publications. In other words, a more well-published author is not more likely to write with a woman.

Interestingly, the results change when including interaction effects with *sex* in the logit estimation, as reported in column 2 of Table 5. The *sex of the first author* is now insignificant, which indicates that women are not in general more likely to have a female coauthor than men are. Also the *mean female share* becomes statistically insignificant, while the *interaction of it with the sex of the first author* is positive and highly significant. This implies that as there is an increase in the share of women in the article's subfields, women's probability of having a female coauthor increases more than men's. More specifically, at the average mean share of women (13 per cent) the marginal effect of an increase of 10 per cent of women in a field, raises the probability of a female author to have a female coauthor with 10 percentage points, while it does not significantly affect the choices of male authors.

	Mean share of women in the field		
	5%	13%	21%
Men	0.23***	0.36	0.61
Women	0.26***	1.01***	3.50*

* denotes significance at least at the 90 per cent level;
** significance at least at the 95 per cent level; and
*** significance at least at the 99 per cent level.

Table 6: Marginal effects on the probability of having a female coauthor of increasing the mean share of women in a field, by gender of first author.

The marginal effect increases with the *mean share of women* in the field for female authors, while it for male authors only has a significant effect when there are very few women in the field, as reported in Table 6. This is in line with the evidence presented in Figure 6, which suggested that men and women react differently to increases in female authors in a field. Generally, the marginal effects of the *mean share of women* are quantitatively more important than all other explanatory variables.

Print year enters positively (and is significant), which means that as time goes by the probability of writing with a woman increases. There is however no gender-specific time trend as the interaction effect with *sex* is insignificant. An article that has *more than two authors* is more likely to have a female coauthor; this probably captures the fact that more authors implies a larger probability that someone of the authors is a woman.

What do these results imply for the predictions of the model? Apparently, the probability that a woman has a female coauthor increases more than that of men as the share of women in the subfields of the article increase. That is, in model terms $\beta_{FTM_m} < \beta_{FTM_f}$. Table 7 summarizes the theoretical implications when men's coauthorship patterns are less sensitive than women's to changes in the gender composition of the field. Hypothesis 1 is clearly not consistent with data since it does not allow for $\beta_{FTM_m} < \beta_{FTM_f}$ so we can start by rejecting it. Two groups of hypotheses remain: on the one hand, hypothesis 2 and 3 which given this first finding predict that the level of single-authoring men is higher than that of single-authoring women, and on the other hand, hypotheses 4, 5, and 6 for which the opposite should hold.

Hypothesis	Female coauthor	Single authored	
1.Gender irrelevance	$\beta_{FTM_m}= \beta_{FTM_f}$	$\alpha_m= \alpha_f$	$\beta_{S_m}= \beta_{S_f}= 0$
2. Gender-specific team preferences, Gender neutrality	$\beta_{FTM_m}< \beta_{FTM_f}$	$\alpha_m> \alpha_f$	$\beta_{S_m}> \beta_{S_f}$
3. Identical team preferences Identical gender preferences	$\beta_{FTM_m}< \beta_{FTM_f}$	$\alpha_m> \alpha_f$	$\beta_{S_m}> \beta_{S_f}$
4. Identical team preferences Symmetric gender preferences	$\beta_{FTM_m}< \beta_{FTM_f}$	$\alpha_m< \alpha_f$	$\beta_{S_m}> \beta_{S_f}$
5. Identical team preferences Female gender preferences	$\beta_{FTM_m}< \beta_{FTM_f}$	$\alpha_m< \alpha_f$	$\beta_{S_m}> \beta_{S_f}$
6.Identical team preferences Male gender preferences	$\beta_{FTM_m}< \beta_{FTM_f}$	$\alpha_m< \alpha_f$	$\beta_{S_m}> \beta_{S_f}$

Table 7: The implications of the model.

We make logit estimations of the probability of writing single-authored articles to test the consistency of the various hypotheses with data. Results are reported in columns 3 and 4 in Table 5 above. Column 3 contains the results from a logit without any interaction effects, which shows that women do not tend to write more single-authored articles everything else equal. The probability of writing single-authored diminishes as the author has an increasing number of publications and as time goes by. When more articles are single-authored in the subfields of the article, the probability of writing single increases, as would be expected. Authors at a *top 9 university* tend to write more single than others.

As we include interaction terms with the sex of the first author, all these effects remain significant. Column 4 shows that women single-author more than men, which implies that $\alpha_f > \alpha_m$ and that we thereby can reject hypothesis 2 and 3. Moreover, the sensitivity of women to the female share in the subfields of the article, is if anything higher than men's, so that $\beta_{S_m} > \beta_{S_f}$ since $\beta_{S_m} < 0$ and $\beta_{S_f} < 0$. Thereby we cannot reject hypotheses 4, 5 or 6. This suggests that there indeed is some kind of gender preferences regarding team formation.

4.4 Robustness, selection and learning

In Table 8 we present some additional logits where we either include new explanatory variables or limit the sample in order to check the robustness of our results. In columns 1 to 3 of Table 8, the dependent variable is the probability of having a female coauthor. In the first column we include an alternative measure for seniority. We use *seniority gap*, that is the gap in number of publications between the first author and the least published coauthor of the article. This variable has no explanatory power indicating there are no specific coauthorship constellations where well-published authors tend to write more with women/men with few publications. Second, we limit the sample to contain only articles that have more than one author. This does not at all alter our results. Finally, we include a dummy for the observations stemming from American Economic Review's Paper and proceedings issue in May every year. Columns 3 and 6 indicate that the May issues of the AER are not driving our results.

	(1)	(2)	(3)	(4)	(5)	(6)
		Probability of having a female coauthor				Pr of single
Sex	9.171 (100.524)	75.387 (113.676)	20.034 (104.504)	-805.4** (235.3)	129.6 (128.3)	182.316** (71.234)
Mean female share	6.824 (4.890)	6.042 (4.990)	6.732 (4.904)	28.59 (19.59)	-1.849 (7.592)	-3.347 (2.694)
Sex * (Mean female share)	12.329*** (4.207)	11.081** (4.392)	11.945*** (4.232)	155.1*** (47.54)	13.94** (6.974)	-5.475* (2.970)
Print year	0.067*** (0.024)	0.055** (0.025)	0.056** (0.025)	0.046 (0.036)	0.061* (0.034)	-0.032** (0.013)
Sex * (Print Year)	-0.005 (0.050)	-0.038 (0.057)	-0.011 (0.052)	0.395*** (0.117)	-0.066 (0.064)	-0.091** (0.036)
Number of publications		0.015 (0.034)	0.046 (0.034)	0.073 (0.048)	0.027 (0.048)	-0.102*** (0.020)
Sex * (N. of publications)		-0.032 (0.100)	0.058 (0.097)	-0.435 (0.266)	0.143 (0.110)	-0.173* (0.096)
Seniority gap	0.017 (0.030)					
Sex * (Seniority gap)	-0.026 (0.108)					
Top 9 university	-0.214 (0.186)	-0.120 (0.208)	-0.331 (0.203)	-0.419 (0.308)	-0.156 (0.268)	0.469*** (0.093)
Sex * (Top 9 university)	0.179 (0.412)	0.408 (0.471)	0.140 (0.439)	2.475*** (0.955)	-0.355 (0.515)	0.408 (0.292)
Mean single						1.877** (0.849)
More than two authors	1.444*** (0.167)	0.707*** (0.163)	1.453*** (0.167)	1.099*** (0.268)	1.688*** (0.220)	
Number of pages	-0.013* (0.008)	-0.021** (0.009)	-0.008 (0.010)	-0.002 (0.013)	-0.021* (0.011)	-0.016*** (0.005)
AER May			0.259 (0.224)			0.295** (0.115)
AER	-0.012 (0.214)	0.099 (0.226)	0.038 (0.227)	0.030 (0.318)	0.015 (0.298)	0.241** (0.121)
JPE	-0.038 (0.265)	0.012 (0.269)	-0.010 (0.272)	-0.131 (0.414)	0.042 (0.357)	0.069 (0.139)
Constant	-138.014*** (47.118)	-113.583** (49.565)	-115.454** (49.369)	-97.471 (71.274)	-123.5* (68.21)	65.799** (25.138)
Observations	3072	1729	3072	1572	1497	3072
Wald Chi^2	160.08	94.29	166.43	68.85	116.00	194.29
Prob> Chi^2	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R^2	0.0954	0.0724	0.0980	0.0895	0.1228	0.0519

Robust standard errors in parentheses. * denotes significance at least at the 90 per cent level; ** significance at least at the 95 per cent level; and *** significance at least at the 99 per cent level. All estimations include dummy variables for JEL-codes (not reported).

Table 8: Some additional logit estimations.

In order to investigate whether there is any indication of a selection process where researchers holding gender segregation preferences select fields where they are more likely to encounter coauthors of the same sex, or if gender preferences change with the gender ratio, we split our data into two groups based on the mean female share in the fields of the article. Since women are in a minority in every field any incentive to select is presumably stronger for them. Hence it is likely that such a selection process would make preferences for gender segregation stronger in fields with relatively more women. An overall weakening of the preferences for segregation as a result of declining inter gender coordination costs would give an effect in the opposite direction. The results, presented in columns 4 and 5, Table 8, show that there are significant differences across the two groups. In the sample of articles with a low ratio of women in its JEL-fields (column 4) the probability of having a female coauthor is much more sensitive to the share of female authors than when there is a high ratio of women (column 5). It is also the case that being a woman by itself reduces the probability of having female coauthors when there are few females to potentially coauthor with. This could indicate that coauthoring an article with a woman is all the more important to women when there are few women in the fields of the article. There is hence no support for any selection of women into fields with a high share of female authors, but rather indications of a higher degree of gender preferences in fields with a lower ratio of women.

5 Concluding discussion

We have modeled the formation of teams as a random matching process influenced by the employees' preferences for team size and gender composition. We have then tested a set of hypotheses on the distributions of gender and team preferences on the pattern of authorships in articles published 1991-2002 in three top economics journals. We find that the gender-gap (female/male) in the propensity of having a female coauthor increases with the proportion of female authorships in the field. This, together with the finding that women single author significantly more than men and that female single authorship declines more than male ditto as the share of female authored articles increases allows us to reject hypotheses of gender irrelevance and gender neutrality. The patterns found in the data conforms with the predictions given by our model when it is assumed that men, women or both, have a form of gender preferences implying that the fraction of the relevant group having a preference for teams with the same sex is larger than the fraction who prefer working with the opposite sex. We can therefore conclude that there is indeed evidence of a preference for gender segregation in team formation in economics at the top of the ability distribution.

References

[1] The Annual Report of the Commission on the Status of Women in the Economics Profession, *American Economic Review*: Papers and Proceedings, several issues.

[2] Becker, Gary S., and Kevin M. Murphy, (1992), "The Division of Labor, Coordination Costs, and Knowledge", *The Quarterly Journal of Economics,* Vol CVII, No 4. pp 1137-1160.

[3] Blanck, Rebecca M., (1991), "The Effects of Double-Blind vsrsus Single-Blind Reviewing: Experimental Evidence from the American Economic Review." *The American Economic Review*, Vol. 81, No. 5, 1041-1067.

[4] Breen, Richard and Garcia-Penalosa, Cecilia, (2002), "Bayesian Learning and Gender Segregation" *Journal of Labor Economics,* vol 20, no 4, pp 899-922.

[5] Durden, G. and P. Gaynor, (1997), "Multi-Authored Papers in Economics", *Journal of Economic Perspectives* 11:2, pp. 91-92.

[6] Ferber, Marianne A. and Michelle Teiman, (1980), "Are Women Economists at a Disadvantage in Publishing Journal Articles?", *Eastern Economic Journal*, vol VI, nos 3-4.

[7] Goyal, Sanjeev, Marco van der Leij and José Luis Moraga-Gonzáles, (2004), "Economics: an emerging small world?" mimeo.

[8] Hamermesh, D. S. and S. M. Oster, (2002), "Tools or Toys? The Impact of High Technology on Scholarly Productivity", *Economic Inquiry* 40:5, pp. 539-555.

[9] Kahn, Shulamit, (1995), "Women in the Economics Profession", *Journal of Economic Perspectives*, vol 9, no. 4. 193-205.

[10] Kahn, Shulamit, (2002), "The Status of Women in Economics during the Nineties: One step forward, Two steps back", Paper presented at AEA Meetings in Atlanta 2002.

[11] Laband, D. N. and R. D. Tollison, (2000), "Intellectual Collaboration", *Journal of Political Economy*, 108:3, pp. 632-662.

[12] Margreiter, Magdalena, Martin G. Kocher, and Matthias Sutter, (2002), "Women in top journals of economics", University of Innsbruck, mimeo.

[13] McDowell, John and Janet Kiholm Smith, (1992), "The effect of gender-sorting on propensity to coauthor: implications for academic promotion", *Economic Inquiry 30*, pp. 68-82.

[14] McDowell, John M., Larry D. Jr Singell and James P. Ziliac, (1999), "Cracks in the Glass Ceiling : Gender and Promotion in the Economics Profession." *American Economic Review, Papers and Proceedings*, vol 89, no.2, pp. 392-396.

[15] McDowell, John M., Larry D. Jr Singell and James P. Ziliac (2001). "Gender and Promotion in the Economics Profession." *Industrial and Labor Relations Review*, vol. 54, no 2.

[16] Prat, Andrea, (2002) "Should a Team Be Homogeneous?" *European Economic Review* (46)7: 1187-1207.

[17] Sauer, R. D., (1988), "Estimates of the Returns to Quality and Coauthorship in Economic Academia", *Journal of Political Economy* 96:4, pp. 855-866.

[18] Tuckman, H. P. and J. Leahey, (1975), "What Is an Article Worth?", *Journal of Political Economy* 83:5, pp. 951-968.

A Summary statistics by journal

Journal	Variable	Obs	Mean	Std. Dev.	Min	Max
AER	*Print year*	1965	1996,27	3,45	1991	2002
	Number of authors	1965	1,68	0,76	1	8
	Female share	1965	0,14	0,30	0	1
	Mixed authorship	1965	0,12	0,32	0	1
	Number of JEL-codes	1965	1,55	0,68	1	4
	Number of pages	1965	9,69	7,94	1	96
JPE	*Print year*	601	1996,33	3,46	1991	2002
	Number of authors	601	1,74	0,74	1	5
	Female share	601	0,10	0,26	0	1
	Mixed authorship	601	0,10	0,30	0	1
	Number of JEL-codes	601	1,69	0,74	0	4
	Number of pages	601	24,95	10,67	1	75
QJE	*Print year*	509	1996,05	3,51	1991	2002
	Number of authors	509	1,82	0,76	1	5
	Female share	509	0,10	0,25	0	1
	Mixed authorship	509	0,10	0,30	0	1
	Number of JEL-codes	509	1,60	0,73	1	5
	Number of pages	509	29,06	10,54	1	62
All	*Print year*	3075	1996,25	3,46	1991	2002
	Number of authors	3075	1,72	0,76	1	8
	Female share	3075	0,12	0,29	0	1
	Mixed authorship	3075	0,11	0,31	0	1
	Number of JEL-codes	3075	1,58	0,70	0	5
	Number of pages	3075	15,88	12,26	1	96

Table A: Summary statistics by journal.

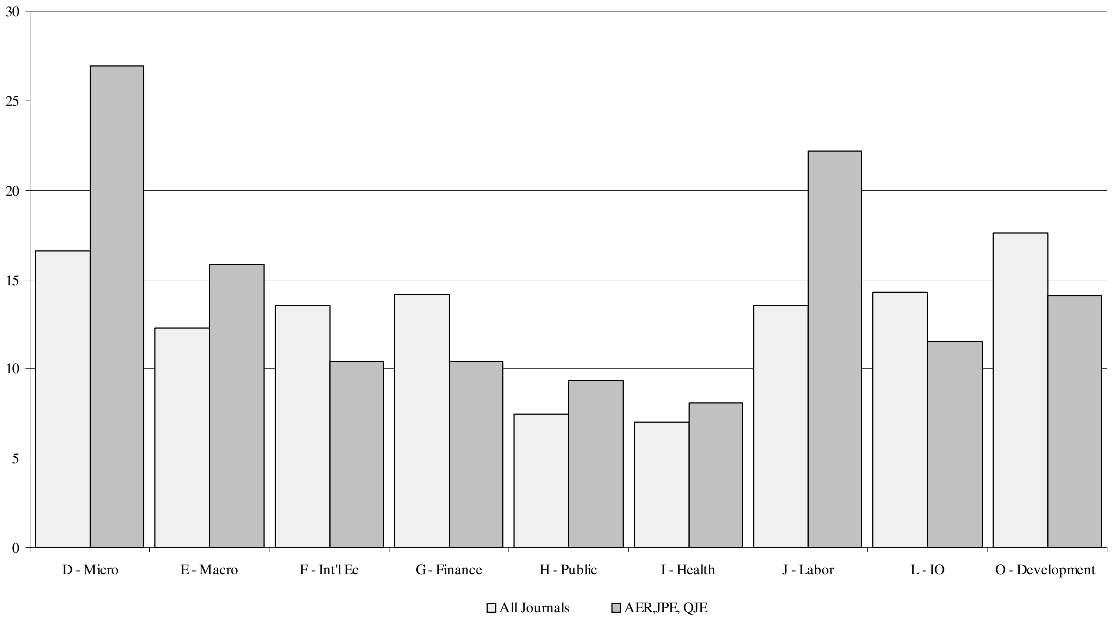

Figure A: The relative frequency of JEL-codes in AER, JPE and QJE compared to all journals in EconLit.

B EconLit classification

The JEL-codes in the data set are those given by EconLit, not the authors. The reason for this is that not all articles contain JEL-codes when they are published. For example, the articles published in the Journal of Political Economy do not have JEL-codes, which means that if we would use only the JEL-codes that appear in the articles we could not include the Journal of Political Economy in our database. We have therefore opted for the JEL-codes reported by EconLit, that employ classifiers to assign JEL-codes and key words to all articles. (This means that there can be a discrepancy between the JEL codes that appear in the publication and those in the EconLit database.)

> "We have no manual on classifications. Our classifiers (graduate and doctoral students of Economics), memorize the classification as they begin to use it. Their classifications are checked, and their misclassifications
>
> are discussed with them. The person in charge of classification (a University Professor of Economics) uses this method to maintain a high degree of consistency (although it can never be perfect). Economics is such a conceptual subject that differences in opinion are common. We try to classify articles under the subject descriptors (up to seven) where we think an economist would look to find such an article. We try to

classify articles under the subject matter, not the theory (such as macroeconomics subdivisions). We put articles under the C, D, and E categories when the subject is the theory."

An advantage of using EconLit's classification is that it is likely to be more homogeneous than if all authors are assigning the JEL codes themselves.

B.1 The JEL classification used in the paper

A - General Economics and Teaching
B - Schools of Economic Thought and Methodology
C - Mathematical and Quantitative Methods
D - Microeconomics
E - Macroeconomics and Monetary Economics
F - International Economics
G - Financial Economics
H - Public Economics
I - Health, Education, and Welfare
J - Labor and Demographic Economics
K - Law and Economics
L - Industrial Organization
M - Business Administration and Business Economics; Marketing; Accounting
N - Economic History
O - Economic Development, Technological Change, and Growth
P - Economic Systems
Q - Agricultural and Natural Resource Economics
R - Urban, Rural, and Regional Economics
Z - Other Special Topics

C Probits

	(1)	(2)	(3)	(4)
	Probability of having a female coauthor		Probability of single-authoring	
Sex	0.449***	-6.401	-0.001	110.742**
	(0.089)	(54.410)	(0.071)	(43.155)
Mean female share	5.437**	3.553	-2.545	-2.040
	(2.393)	(2.401)	(1.608)	(1.653)
Sex * (Mean female share)		7.159***		-3.083*
		(2.184)		(1.783)
Print year	0.028**	0.028**	-0.025***	-0.018**
	(0.011)	(0.012)	(0.007)	(0.008)
Sex * (Print Year)		0.003		-0.055**
		(0.027)		(0.022)
Number of publications	0.027*	0.027	-0.063***	-0.060***
	(0.016)	(0.017)	(0.012)	(0.012)
Sex * (N. of publications)		0.024		-0.106*
		(0.055)		(0.055)
Top 9 university	-0.138	-0.150	0.328***	0.313***
	(0.086)	(0.093)	(0.053)	(0.056)
Sex * (Top 9 university)		0.066		0.230
		(0.226)		(0.173)
Mean single			1.398***	1.244**
			(0.514)	(0.520)
More than two authors	0.743***	0.750***		
	(0.087)	(0.087)		
Number of pages	-0.006	-0.006	-0.013***	-0.014***
	(0.004)	(0.004)	(0.003)	(0.003)
AER	0.004	0.004	0.113	0.113
	(0.105)	(0.105)	(0.074)	(0.074)
JPE	-0.018	-0.028	0.007	0.013
	(0.128)	(0.128)	(0.085)	(0.085)
Constant	-58.441***	-58.192**	49.267***	34.819**
	(21.779)	(23.827)	(14.466)	(15.439)
Observations	3072	3072	3072	3072
Wald Chi^2	142.98	160.30	182.91	199.77
Prob> Chi^2	0.000	0.000	0.000	0.000
Pseudo R^2	0.0915	0.0980	0.0462	0.0502

Robust standard errors in parentheses. * denotes significance at least at the 90 per cent level; ** significance at least at the 95 per cent level; and *** significance at least at the 99 per cent level. All estimations include dummy variables for JEL-codes (not reported).

Table C: Binomial probits for the basic specifications.