Gender Discrimination and Evaluators' Gender:

Evidence from the Italian Academia

03/07/2015

Maria De Paola, Vincenzo Scoppa*

Department of Economics and Statistics, University of Calabria, and IZA (Bonn)

Relying on a natural experiment consisting in 130 competitions for promotion to associate and full professor in the Italian University, we analyze whether gender discrimination is affected by the gender of evaluators. Taking advantage of the random assignment of evaluators to each competition, we examine the probability of success of each candidate in relation to the committee gender composition, controlling for candidates' scientific productivity and a number of individual characteristics. We find that female candidates are less likely to be promoted when the committee is composed exclusively by males, while the gender gap disappears when the candidates are evaluated by a mixed sex committee. Results are qualitatively similar across fields and type of competitions and are robust to the exclusion of candidates who have withdrawn from competition and when controlling for a number of evaluators' characteristics.

JEL classification: J71; M51; J45; J16; D72, D78

Keywords: Gender Discrimination; Evaluators' Gender; Affirmative Actions; Academic Promotions; Withdrawal Decision; Natural Experiment; Random Assignment.

Introduction

Female educational levels and female labor force participation have recently risen in most countries. Nonetheless, in many spheres of social and economic life gender inequality is still pervasive. A huge literature shows that female employees earn less than males even when they have the same levels of education, work experience and professional qualification (see, among others, Blau and Kahn 2003; Altonji and Blank 1999; Weichselbaumer and Winter-Ebmer 2005). A number of papers show that in many countries the gender wage gap is increasing across the wage distribution (Arulampalam *et al.* 2007;

* Department of Economics, Statistics and Finance, University of Calabria, 87036 Arcavacata di Rende (CS), Italy and IZA (Bonn). E-mail: m.depaola@unical.it; v.scoppa@unical.it. We have benefited from helpful comments from the Editor Peter Norman Sorensen and two anonymous referees. We also would like to thank Manuel Bagues, Giorgio Brunello, Lorenzo Cappellari, Alessandra Casarico, Giovanni Fattori, Davide Fiaschi, Margherita Fort, Anna Giunta, Luca Gori, Andrea Ichino, Myriam Mariani, Patrizia Ordine, Michela Ponzo, Paola Profeta, Giuseppe Rose, Manuela Stranges, Eliana Viviano and seminar participants to the European Association of Labour Economists (EALE), Bonn, September 2012, the Italian Association of Labour Economists (AIEL), Milano 2011, Universities of Padova, Bologna, Milano-Bocconi.

Albrecht *et al.* 2003) and that women face the so called "glass ceiling", that is, they remain greatly underrepresented in higher paying jobs and in top positions. These results can be due to undersized females' investments in human capital or less experience, but they may also be related to the fact that promotion procedures favor men rather than women. For example, some recent works examining promotions and pay in the academic labor market show that women suffer a disadvantage in promotions and a within-rank pay gap (Blackaby *et al.* 2005; McDowell *et al.* 1999; Ginther and Kahn 2004).

The economic theory has tried to explain the male-female gap in labor market outcomes (not imputable to skill differences) considering three main channels (see Altonji and Blank 1999, for a review). The first focuses on tastes and is based on the idea that some decision-makers (the employer, other workers, clients) dislike to interact with females (Becker 1957). The second is, instead, based on incomplete information: statistical discrimination arises because employers possess limited information about skills or productivity of candidates and use easily observable characteristics to infer productivity and so negative prior beliefs about the abilities of some groups may become self-fulfilling (Aigner and Cain 1977). Finally, some recent works have pointed to differences in preferences and psychological attitudes between males and females: less competitive behaviors, greater risk aversion and less bargaining attitudes may be partly responsible for worse females' labor market outcomes (Bertrand 2011).

An interesting issue is to what extent discrimination depends on the gender of evaluators. At the best of our knowledge only a few works have tried to examine this issue. Two recent papers by Bagues and Esteve-Volart (2010) and by Zinovyeva and Bagues (2011) based, respectively, on a recruitment procedure for positions in the Spanish Judiciary and on competitions to associate and full professor positions in Spain, reach rather ambiguous results. Whereas from Bagues and Esteve-Volart (2010) it emerges that female candidates are less likely to be hired when the randomly assigned selection committee is characterized by a higher percentage of female evaluators, Zinovyeva and Bagues (2011) show that committees with a relatively larger share of females reduce gender discrimination against women in competitions to full professors positions, but they find no statistically significant effect as regards competitions to associate professor.

A different approach has been taken by some works investigating the impact produced by female bosses on female employees. Bell (2005) using US data finds that women-led firms hire more top

executive women and pay higher wages to female workers compared to men-led firms. Similarly, Cardoso and Winter-Ebmer (2007), using data from Portugal, show that female leadership in firms leads to higher wages for women and lower wages for males. Goldin and Rouse (2000) show that female musicians have increased their probability of being hired in prevailingly male symphony orchestra after the adoption of "blind" auditions with a "screen" to conceal the candidate's identity from the jury. In an educational context, Lavy (2008), analyzing the existence of gender discrimination in teachers' evaluation of students comparing results in a blind and in a non-blind test, shows that the gender bias is sensitive to the gender of evaluators, but the direction of the effect varies across disciplines.

In this paper we try to shed more light on this issue providing new evidence on whether the gender of evaluators matters for discrimination. We base our analysis on a natural experiment involving the Italian academic promotion system for associate and full professor positions. Our framework shares with the papers of Bagues and Esteve-Volart (2010) and Zinovyeva and Bagues (2011) the same identification strategy, based on the random assignment of evaluators to competitions.

Thanks to the random assignment procedure followed in Italy to select the members of evaluation committees for competitions to associate and full professor positions opened in 2008, we are able to estimate the probability of success of candidates in relation to the committee gender composition, avoiding endogeneity problems deriving from unobservable factors that may be correlated with committees' and candidates' characteristics.

Unfortunately, data on these competitions are not readily available and have to be collected reading the official reports produced by each committee. Since collecting data on all promotion procedures would have been an unmanageable task, we have decided to focus on promotion procedures in only two fields: Economics and Chemistry. More precisely, we use data on 130 public competitions involving about 1,000 candidates evaluated by 650 professors.

For each committee member and for each candidate we have collected data on the number of publications, the number of citations, *h* and *g* indexes, and on the university where they worked at the time of the competition. We have used these information to build indicators of candidates' and committees' scientific productivity, to identify whether candidates are insiders in the university opening the vacancy and to find out professional networks between candidates and committee members.

Controlling for all these factors, we have estimated the probability of success of each candidate. From our analysis it emerges that female candidates have a lower probability of success compared to their male counterparts (3.7 percentage points less). We also find that female candidates are significantly less likely to be promoted when the randomly assigned committee is composed exclusively by males: in this case the probability of success of females is about 6-7 percentage points less. On the contrary, the presence of female members in the committee allows to overcome almost completely the discrimination against women. This result holds true both for the Economics and Chemistry fields and is robust also when we exclude from our sample the candidates who have withdrawn from competition. As regards heterogeneous effects across different type of positions, we find that in competitions to associate professor, committees composed exclusively by males operate a stronger discrimination against women with respect to that emerging in competitions to full professor positions. Moreover, in competitions to associate professor, the improvement in female outcomes produced by a mixed sex committee is smaller in magnitude.

Our work contributes to the literature analyzing the nature of gender discrimination in high-paying jobs and top positions. We document the persistence of a gender gap in promotions even after controlling for relatively good measures of productivity. Thanks to the information we have on a large number of individual characteristics and on a number of quite reliable measures of individual productivity, we are confident that – in comparison to the large part of the literature on gender wage gap (see Cahuc and Zylberberg 2004) – our results are less affected by problems deriving from unobservable characteristics, unbalanced across gender, that may determine individual earnings.

We also add to the small literature analyzing the evaluators' gender effect. Since results reached by the only two existing papers analyzing this issue are mixed, we think it is very useful to provide new evidence.

The paper is organized as follows. Section I presents the Italian academic promotion system and describes the data used in our analysis. In section II we carry out some random assignment checks. In section III we show our estimation results on the impact of committee gender composition on female candidates' probability of success. Section IV is devoted at investigating differences across types of position and fields. In section V we offer a set of robustness checks. Section VI concludes.

I. Institutional Background and Data

Italy is one of the worst performing countries in terms of gender equality: in 2012, the Gender Gap Index ranks Italy at the 80th position. Women are underrepresented both in the public and in the private sector. Only 20% of seats in the parliament are held by women and only 3% of the 50 largest companies' board directors are women. As far as the academia is concerned, women account for 45% of assistant professors, 34% of associate professors and for 20% of full professors. Although the number of women in the lower ranks has grown over time, the increase has been modest among higher positions.[1]

The rules governing careers in the Italian Universities have changed over time. Abandoning a centralized and nationwide competition, a new mechanism was implemented for promotion to associate and full professor positions since 1999: each university willing to fill a vacancy initiated a competition and a committee of five members was selected to choose two or three winners (the so called "idonei"). One member of the committee was appointed by the university opening the vacancy and the remaining four were elected by all professors in the field.

These rules were strongly criticized because elected committee members were not typically chosen with the aim to screen the best candidates but according to agreements among influential members of the academia, with the result that promotions were far from being related to candidates' scientific productivity.[2] Nevertheless, in 2008, under this system, a huge number of vacancies (695 positions for full professors and 1,110 positions for associate professors) were opened by Italian Universities. At the end of 2008, the Italian Government, worried of the outcomes that could arise by the system in force, has decided to change the rules governing promotions to associate and full professor positions. The main change has concerned the way in which committees are selected: it has been established that four members out of five

[1] Similar and even worse figures can be found for other countries. For example, in UK universities women made up 36% of professors (Higher Education Statistics Agency, 2010). In US, in 2005, about 30% of assistant professors in economics were women, while the share of women among associate and full professors in the same field was of 15.6 percent (see Ginther and Kahn 2004).

[2] Analysing the working of the Italian academic competitions, Perotti (2002) describes the system as follows: "University X wants to promote its own insider, and initiates a competition. The commissioner from university Y supports "idoneità" [promotion] for the insider of university X, with the mutual understanding that university X will return the favour in the future when it comes to promoting university Y's insider".

have to be randomly selected (among all the full professors in each field)[3] instead of being elected, while, as in the previous system, one member is appointed by the university opening the vacancy. The internal member is selected before the other members of the committee are randomly selected.

The purpose of the reform was to increase the independence of the external members and to diminishing the influence of the internal member. The internal member is typically supposed to support the candidate preferred by his/her university. This preference can be unrelated to the quality of candidates as the Italian university systems lacks of a mechanism linking significantly funding to performance in research.

In the Appendix A we provide a simple econometric analysis to evaluate if the sheer change in the system has produced any effect on women's promotion chances. We find that women's prospects have significantly improved under the new system.

Following the new system committee members meet to evaluate candidates and at the end of the evaluation process two winners for each evaluation procedure are selected. In competitions to full professor candidates are evaluated exclusively on the basis of their CV and there are no interactions between committees' members and candidates. In competitions to associate professor skills shown by candidates in a teaching lecture are also taken into account. In addition, candidates have to present and discuss with the evaluation committee the methodology and the results obtained in their research activity. In both types of competitions evaluators are full professors. As in the previous system, the University that has initiated the competition can decide to appoint one of the winning candidates as professor, while the other can be appointed by another university within three years.

As explained above, data on competitions have to be collected reading the final report produced by each committee at the end of the evaluation process. Due to the huge amount of work related to data collection, we have chosen to focus our attention exclusively on competitions undertaken in two relatively large fields: Economics (5 sub-fields) and Chemistry (10 sub-fields).[4] We have chosen these two fields with the aim of analyzing both a scientific and a social science field. Among scientific fields, Chemistry

[3] The selection is carried out by the officials of the Ministry of Education, University and Research, through a computerized random procedure certified by a notary.

[4] In Economics, 28% of professors are females (women account for 42% of assistant professors, 26% of associate professors and 16% of full professors). In Chemistry, 42% of professors are females (women account for 57% of assistant professors, 40% of associate professors and 18% of full professors).

was characterized by a quite large proportion of females, while other possible fields, such as Physics or Engineering were excluded due to the extremely small number of female evaluators (mirroring the scarce presence of females in the field). Among social science fields, we have focused on Economics because it was easier to find measures of individual productivity compared for example to Humanities or Sociology.

Competitions are opened in each sub-field. In Economics there are five sub-fields (Economics, Political Economy, Econometrics, Public Economics, Applied Economics) but sub-fields are not very specific and it is frequent that a professor in Economics obtains a promotion in Political Economy and viceversa. For Chemistry, instead, sub-fields such as Organic Chemistry, Inorganic Chemistry, Physical Chemistry etc. are somehow more specific.

By February 2011, 52 competitions (31 to associate professor and 21 to full professor) were concluded in the Economics field, while in Chemistry 78 competitions (46 to associate professor and 32 to full professor) were completed. As a consequence, we end up with 130 evaluation procedures, involving 1,007 candidates and 650 committee members. The average number of competitors for each competition is equal to 17.53. Candidates were allowed to apply to a maximum of 5 different competitions. Each candidate has applied on average to 2 competitions. The total number of observations at the candidate-competition level is equal to 2,279.

During the evaluation process about 27% of candidates decided to withdraw from competition. Withdrawals are more frequent in competitions to associate professor positions (43.6%) than in competition to full professor positions (7.8%). The sample including only the candidates that maintain their candidacy until the conclusion of the evaluation procedure is made of 1,652 observations.

We have collected the list of evaluators, candidates and winners from the final reports produced by each committee. The gender has been inferred from the first name. Age has been taken from official reports or searching CVs on-line. In the few cases in which we were not able to find the year of birth we have imputed it as the year of graduation minus 24 (the age at which typically high ability students graduate).

To gather information on the scientific productivity of candidates and evaluators we have used the "Publish or Perish" software based on Google Scholar. More precisely, we have collected data on the

number of publications, citations, *h* and *g* indexes,[5] for each individual at the date of conclusion of each competition.[6] We have decided to consider the publications until this date instead of until the date of application since long delays typically occur from when papers are accepted for publication (and candidates include them in their CVs) and when publications appear as published in scientific journals.

Using data on the number of publications and citations and on the *h* and *g* indexes, we have undertaken a principal component analysis to obtain a comprehensive measure of individual productivity (only the first component is considered), which we call *Productivity*. For each candidate we build *Relative Productivity* as the difference between his/her *Productivity* minus the average productivity of the other candidates in the competition. Moreover, this measure of productivity is used to calculate for each evaluation committee the average productivity of evaluators, considering only the four randomly selected members.

The affiliations of both evaluators and candidates have been obtained from the Ministry of Education, University and Research (MIUR)[7] and used to build a dummy variable *Insider* taking the value of one for candidates who work in the university opening the vacancy. Moreover, we build an indicator of professional networks between candidates and committee members, *Connections*, taking the value of one when there is at least a committee member (excluding the internal appointed evaluator) from the same university as the candidate and zero otherwise.

Descriptive statistics for candidates and for evaluators are reported in Table 1. The percentage of female candidates is about 40%, higher in competitions to associate professors (45%) than in competitions to full professors (33%). Candidates to full professor positions over their lifetime have published on average 61 works receiving 469 citations, whereas the average number of publications of candidates to associate professor was 41 with 274 citations. About 15% of candidates are insiders and 10% of them has connections with at least one member of the committee. The great majority of candidates is performing an academic job (90%). On average, candidates are 44.7 years old, candidates to associate professor positions

[5] The *h* index (Hirsch index) is a measure of both the productivity and the impact of published works (based on citations received) of a researcher. A scientist has index *h* if *h* of his/her *N* papers have at least *h* citations each, and the other (*N*− *h*) papers have no more than *h* citations each. The *g* index is defined in a similar way but gives higher weight to highly cited paper. More precisely, given a set of articles ranked in decreasing order of the number of citations that they received, the *g* index is the largest number such that the top *g* articles received (together) at least g^2 citations.

[6] For Economics we also consider the Impact Factor of the Journals in which candidates publish (see Section 5).

[7] From the web page: http://cercauniversita.cineca.it/php5/docenti/cerca.php

are typically younger (41.7) than candidates to full professor positions (48.3) and Chemistry candidates are older (46.7) than Economics ones (42.1).

As regards evaluators, we focus our attention exclusively on the four randomly selected committee members and neglect the internal member since the individual characteristics of the latter could be correlated to unobservable determinants of success of candidates. About 16% of the randomly selected evaluators are females. 55% of committees are composed exclusively by males, 31% has one female member, and 10% and 4% percent of committees has, respectively, 2 and 3 female members. Given this distribution, we build a dummy variable *Females in Committee* taking the value of one when at least one female was among the committee members: 44.6% of committees have among their members a female evaluator. The average age of evaluators is 60 and about 28% of evaluators are from Universities of the South of Italy. On average committee members over their lifetime have published 82 papers receiving 779 citations.

Table 1. Descriptive Statistics

	Mean	*St. Dev*	*Min*	*Max*	*Observations*	*Median*
		Candidates				
Success	0.113	0.317	0	1	2279	
Associate Professor	0.122	0.327	0	1	1024	
Full Professor	0.103	0.304	0	1	1255	
Female	0.397	0.489	0	1	2279	
Associate Professor	0.453	0.498	0	1	1024	
Full Professor	0.328	0.469	0	1	1255	
Number of Papers	50.491	37.740	0	269	2279	41
Associate Professor	41.909	30.080	0	199	1024	51
Full Professor	61.010	43.144	0	269	1255	35
Citations	362.502	491.504	0	4485	2279	219
Associate Professor	274.811	384.017	0	4485	1024	182
Full Professor	469.976	579.806	0	4431	1255	313
h-index	8.812	5.404	0	36	2279	8
Associate Professor	7.726	4.814	0	36	1024	9.5
Full Professor	10.144	5.779	0	35	1255	7
g-index	14.646	9.409	0	65	2279	14
Associate Professor	12.778	8.348	0	64	1255	12
Full Professor	16.935	10.109	0	65	1024	16
Relative Productivity	0	1.758	-5.256	10.593	2279	
Associate Professor	0	2.095	-5.256	9.387	1024	
Full Professor	0	1.426	-5.081	10.593	1255	
Insider	0.147	0.355	0	1	2279	
Associate Professor	0.168	0.329	0	1	1024	
Full Professor	0.123	0.374	0	1	1255	
Connections	0.103	0.305	0	1	2279	
Associate Professor	0.112	0.317	0	1	1024	
Full Professor	0.091	0.289	0	1	1255	
Age	44.697	6.771	29	69	2279	
Associate Professor	41.728	5.330	30	69	1024	
Full Professor	48.337	6.572	29	67	1255	
University Job	0.902	0.298	0	1	2279	
Associate Professor	0.873	0.333	0	1	1024	
Full Professor	0.936	0.244	0	1	1255	
Withdrawn	0.275	0.446	0	1	2279	
Associate Professor	0.436	0.496	0	1	1024	
Full Professor	0.078	0.268	0	1	1255	
		Committees' Members				
Females in Committee	0.446	0.4990	0	1	130	
% Females in Committee	0.156	0.2047	0	0.75	130	
Age Com. members	60.480	3.5786	49.5	69	130	
Perc. of Com. members South	0.283	0.2459	0	1	130	
Number of papers Com. members	82.943	88.7533	1	819	520	
Number citations Com. members	779.916	1591.934	0	22370	520	
Associate Position	0.592	0.4933	0	1	130	
Economics	0.400	0.4918	0	1	130	

In Table 2 we report descriptive statistics for candidates and for evaluators separated by gender. Female candidates have a lower probability of success than males and are characterized by lower measures of productivity. We have tested whether pre-determined characteristics differ by gender, regressing each variable, in turn, on *Female*, on a dummy for type of position and sub-field dummies. We find that males perform significantly better in variables measuring productivity (number of papers,

citations, *h*, *g*) whereas, as regards other characteristics, we do not find any significant difference by gender. Thus, to avoid estimation biases in our regressions, it is necessary to control for measures of candidates' productivity.

As far as committee members are concerned, male and female evaluators have almost the same age, but seem to differ both in terms of productivity, with males showing a higher scientific productivity, and in terms of the geographical area in which they work (about 28% of male evaluators are employed in a University located in the South, while this figure is 25% for female evaluators). We have tested the statistical significance of these difference and in all cases (with the exception of working in the South, which is statistically significant with a *p*-value of 0.08) we are not able to reject the null hypothesis of zero differences.

Table 2. Descriptive Statistics by Gender

	Mean	*St. Dev*	*Observations*	*Mean*	*St. Dev*	*Observations*
Candidates						
	Females			*Males*		
Success	0.089	0.286	905	0.130	0.337	1374
Number of Papers	43.791	32.002	905	54.905	40.494	1374
Citations	323.314	477.636	905	388.315	498.922	1374
h-index	8.454	4.990	905	9.049	5.649	1374
g-index	14.009	8.875	905	15.065	9.726	1374
Relative Productivity	-0.183	1.587	905	0.120	1.854	1374
Insider	0.159	0.366	905	0.140	0.347	1374
Connections	0.106	0.308	905	0.102	0.303	1374
Age	44.927	5.999	905	44.547	7.233	1374
University Job	0.908	0.289	905	0.897	0.304	1374
Withdrawn	0.306	0.461	905	0.255	0.436	1374
Committees' members						
	Females			*Males*		
Age	59.361	5.674	81	60.687	6.806	439
Area of work South	0.235	0.232	81	0.286	0.246	439
Number of papers	71.013	60.118	81	81.727	84.885	439
Number citations	626.235	814.556	81	705.223	1173.077	439
h-index	11.197	7.664	81	11.210	8.000	439
g-index	18.605	13.000	81	18.747	14.127	439

II. Random Assignment Checks

Our identification strategy is based on the random assignment of committee members to each competition. To investigate the randomness of the assignment mechanism, we regress a number of individual

characteristics of candidates participating at each competition (the percentage of female candidates, the average productivity of candidates, the number of candidates competing for the position, the percentage of insider candidates) on the dummy *Females in Committee*, controlling for sub-field dummies (since the random assignment of evaluators to competitions was conditional on sub-fields) and for the type of position (a dummy taking value of one for competitions to associate positions and zero otherwise). Moreover, since in building *Females in Committee* we only consider the percentage of females among the randomly selected evaluators without considering the committee member appointed by the university opening the vacancy, we also check for any possible correlation between the characteristics of the internal member and the presence of females among the randomly selected committee members.

The estimation results are reported in Table 3. In column (1) we show that the correlation between the percentage of female candidates and the presence of females evaluators is far from being statistically significant. Similarly, there is no statistically significant correlation between the presence of females in the committee and the average productivity of candidates (column 2), the number of competing candidates (column 3) and the percentage of insider candidates (column 4). In columns 5 and 6 we show that the presence of female evaluators in the committee is not related to the gender or to the scientific productivity of the internal commissioner.[8]

We also checked whether the predetermined characteristics are related to the percentage of females in each committee (*% Females in Committee*) instead of using the dummy *Females in Committee*, obtaining very similar results (not reported).

As an alternative check, we have used our measures of committee gender composition as dependent variables, regressing them on the full set of variables describing predetermined characteristics at the competition level (percentage of female candidates, candidates' average productivity, number of competitors, percentage of insiders) and have tested for the joint significance of these covariates. Results (not reported) do not allow us to reject the null hypothesis of zero effects (the *F*-test is equal to 0.67 with a *p*-value of 0.68).

[8] This high coefficient for the productivity of the internal commissioner is due to some outliers. The average productivity of the internal commissioner is 0.26, but it has a high standard deviation because of the influence of two outliers (with a productivity of 11.4 and 13.2). If we re-estimate specification (6) excluding these two observations, the coefficient reduces to 0.102 and the corresponding *p*-value is equal to 0.73.

Since female committee members were not sorted into competitions according to the characteristics of the candidates or of the internal commissioner, we conclude that the assignment of evaluators to each competition has been effectively random and, as a consequence, any effect of the committee composition can be interpreted as causal.

Table 3. Regressions for Random Assignment Checks

	(1) % Female Candidates	(2) Candidates' Average Productivity	(3) Number of Competitors	(4) % Insiders	(5) Female Internal Commissioner	(6) Productivity Internal Commissioner
Females in Committee	0.002	0.064	0.804	0.014	0.004	-0.428
	(0.026)	(0.114)	(1.262)	(0.023)	(0.060)	(0.381)
Observations	130	130	130	130	130	130

Data at competition level. The dependent variable is reported at the top of each column. Standard errors are reported in parenthesis. In all the regressions we control for sub-field dummies and for type of position dummy.

III. Gender Discrimination and Evaluators' Gender: The Empirical Findings

To uncover the effect of committee gender composition on the probability of success of candidates we estimate the following model:

[1] $$Success_{ij} = \beta_0 + \beta_1 Female_i + \beta_2 Females\ in\ Committee_j + \beta_3 Female_i * Females\ in\ Committee_j + \phi X_{ij} + \mu_j + \lambda_j + \varepsilon_{ij}$$

where the dummy variable $Success_{ij}$, taking value of 1 if candidate *i* has won competition *j*, depends on the candidate gender $Female_i$, on a vector X_{ij} of the candidate characteristics (including scientific productivity and the dummies $Insider_{ij}$ and $Connections_{ij}$), our indicator of committee gender composition, $Females\ in\ Committee_j$, a dummy for the type of position, λ_j, and dummies for scientific sub-fields μ_j. To investigate whether the probability of success of candidates is affected by the gender composition of the committee, we include among our regressors the interaction term between $Female_i$ and $Females\ in\ Committee_j$. Therefore, the coefficient β_1 measures the effect of being a female on the probability of success when the evaluation committee is composed exclusively by men, while $\beta_1 + \beta_3$ represents the extent of female discrimination (if any) when there is at least a female among the committee members.

Equation [1] is estimated using a probit model. Marginal effects are reported in Table 4. We base our analysis on the whole sample of candidates applying for a position to associate or full professor, independently from their effective participation in the competition (in Section 6.1 we exclude candidates withdrawing from competition). In all the regressions, to take into account common shocks that may affect the performance of all candidates participating to the competition, standard errors are allowed for correlation at the competition level.

In the first specification of Table 4 (column 1) we estimate the difference in the probability of success between males and females controlling for scientific sub-field dummies, type of position and number of competing candidates, without taking into account other candidates' characteristics: a female has a lower probability of success of about 4.7 percentage points (significant at the 1 percent level). Since on average the probability of success for a male is about 13%, females suffer a reduction of about 36% in the chances of winning a competition.

In column 2, in order to avoid the bias that may derive from the fact that candidate's gender may be related to some individual features affecting the probability of success, we include among regressors the comprehensive measure of scientific productivity, *Relative Productivity*, and the dummy variables *Insider*, *Connections* and *University Job*. The candidate's scientific productivity contributes to the probability of winning though the effect is quite small in magnitude: an increase of one standard deviation in *Relative Productivity* produces an increase in the probability of success of about 3.5 percentage points. On the other hand, it emerges that being an insider strongly improves the probability of success (by 28 percentage points). *Connections* are also relevant and increase the probability of success by 7 percentage points. The dummy *University Job* is not statistically significant. Importantly, also controlling for these characteristics, it emerges that females suffer a reduction of 3.7 percentage points in the probability of success.

In column 3 we estimate specification 1 adding as regressors the dummy $Females in Committee_j$ and the interaction term $Female_i * Females\ in\ Committee_j$. It emerges that the presence of female members in the committee increases the probability of success of female candidates. More precisely, all-males committees reduce the probability of success of female candidates by 7.6 percentage points, while mixed sex committees eliminates the gender discrimination against women. When among the committee

members is present at least a female, the difference between males and females in the probability of success turns out to be equal to -0.014 with a *p*-value of 0.489.[9]

To analyse whether the gender discrimination implemented by all-males committees is related to individual characteristics, in column 4 we include the controls considered in column 2. Our results remain substantially unchanged as we find that committees composed exclusively by men discriminate against women, reducing their probability of success by about 6.4 percentage points. However, gender discrimination disappears when candidates are judged by a mixed sex committee: the difference between males and females is equal to -0.007 (*p*-value =0.737).

Since we are controlling for scientific productivity, the estimated gender difference cannot be imputed to any difference of productivity between males and females. Moreover, given the controls for *Insider* and *Connections*, we are able to exclude that the uncovered effect depends on possible differences between males and females in the probability of being an insider or having a connection.

To have an idea of the magnitude of the effect produced by a mixed sex committee on female candidates' probability of success, consider that it is equivalent to the improvement deriving from an increase of 2 standard deviations in a candidate's *Relative Productivity*.

These results hold true also when – instead of using *Relative Productivity* – we consider separately our different measures of individual scientific productivity. In column 5 we report results obtained using the relative *h* index. Similar findings are obtained also using, alternatively, the number of publications, citations or the *g* index (not reported).

[9] As in probit models the interaction effect is the cross-partial derivative of the expected value of the dependent variable, the interaction effect cannot be interpreted straightforwardly. To investigate the effect of mixed sex committee on female candidates' probability of success we have used the Stata command *predictnl*.

Table 4. Probit Estimates of the Probability of Success

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.047***	-0.037***	-0.076***	-0.064***	-0.065***	-0.051***	-0.063***
	(0.014)	(0.012)	(0.017)	(0.016)	(0.016)	(0.015)	(0.016)
Female*(Females in Com.)			0.076**	0.072*	0.073*		0.069**
			(0.037)	(0.037)	(0.036)		(0.036)
Female*(% Females in Com.)						0.089	
						(0.063)	
Females in Com.			-0.021**	-0.020*	-0.020**		-0.017**
			(0.010)	(0.010)	(0.010)		(0.011)
% Females in Com.						-0.023	
						(0.026)	
Relative Productivity		0.020***		0.020***		0.020***	0.022***
		(0.003)		(0.003)		(0.003)	(0.004)
Insider		0.283***		0.283***	0.282***	0.283***	0.283***
		(0.029)		(0.029)	(0.029)	(0.029)	(0.029)
Connections		0.070***		0.071***	0.069***	0.070***	0.071***
		(0.026)		(0.026)	(0.026)	(0.026)	(0.026)
University Job		-0.007		-0.006	-0.011	-0.007	-0.006
		(0.019)		(0.020)	(0.020)	(0.020)	(0.020)
Age		-0.001		-0.001	-0.001	-0.001	-0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Relative h-index					0.008***		
					(0.001)		
(Relative Prod.)* (Females in Com.)							-0.005
							(0.006)
Relative Prod.)* (Females in Com.)*Female							0.004
							(0.006)
Observations	2279	2279	2279	2279	2279	2279	2279
Pseudo R-squared	0.041	0.166	0.066	0.170	0.167	0.167	0.170

Notes: The Table reports marginal effects of Probit estimates (evaluated at the mean values of the explanatory variables in the sample). The dependent variable is *Success*. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

In column 6, instead of the dummy variable $Females\ in\ Committee_j$, we consider the *% Females in* $Committee_j$ among the randomly selected members of the committee as a measure of committee sex composition. From the coefficient on the interaction (*Female*)*(*% Females in* $Committee_j$) it turns out that one more female in the committee increases the probability of success of female candidates of 2.2 percentage points. The presence of two female members in the committee reduces to zero the bias against women.

In our analysis it is difficult to have reliable estimates for the presences of more than one female member in the evaluation committee since, unfortunately, only in 18 competitions (out of 130) the number of females in committees is greater than one. We have tried to investigate this issue by including among regressors a dummy for *One Female in Committee*, a dummy *Two or more Females in Committee* and two interactions terms *Female*(One Female in Committee)* and *Female*(Two or more Females in Committee)*. Similarly to the previous findings, when there is one female in the committee, discrimination

disappears: the test for *Female+Female*(One Female in Committee)* has a p-value of 0.88. However, we find a rather unclear result as regards the interaction *Female*(Two or more Females in Committee)*. The coefficient is lower in magnitude with respect to *Female*(One Female in Committee)* (+0.042 instead of +0.067) but we cannot reject neither the hypothesis that the two interactions are equal (p-value=0.549) nor the hypothesis that *Female*(Two or more Females in Com.)* is different from zero (p-value=0.26). On the other hand, discrimination seems to disappear also when there are two or more females in committees: the test for *Female+Female*(Two or more Females in Com.)* has a p-value of 0.38. These ambiguous results are probably caused by the fact that in our sample competitions with two or more females are a very small number and there is not enough statistical power.

Finally, in column (7) we investigate whether evaluation committees with female members are more likely to select productive candidates. At this aim, we build two interaction variables *Relative Productivity*Female in Committee* and *Relative Productivity*Female in Committee*Female* and estimate the specification (4) adding these interaction terms among regressors. From our estimates, we see that the interactions are very far from statistical significance: we do not find any significant difference in the importance of productivity for success (among committees with and without female members and among female and male candidates). Therefore, the effect of mixed gender committees in favour of female candidates has not changed (neither in positive nor in negative) the average "quality" level of successful candidates.

In Figure 1 we plot the probability of success of male and female candidates, separately for all-males and mixed sex committees, in relation to their relative scientific productivity based on the results of column 4 (Table 4). The vertical distance between the continuous line (above in the figure) and the dashed line (below in the figure) represents the gender discrimination when the judging committee is composed exclusively by males. As shown in the Figure, this gap tends to close when the judging committee is composed also by female members: the two lines representing male's (line with dashes and dots) and female's (dotted line) probability of success with a mixed sex committee are very close to each other.[10]

[10] The gender gap seems to be larger when relative productivity is also larger. However, this effect is probably due to the fact that candidates with a low relative-productivity rarely win the competition. Then, for candidates with low relative-productivity gender discrimination is not particularly relevant: regardless to their gender, candidates with low productivity have a low probability to win a competition. On the contrary, when relative-productivity increases

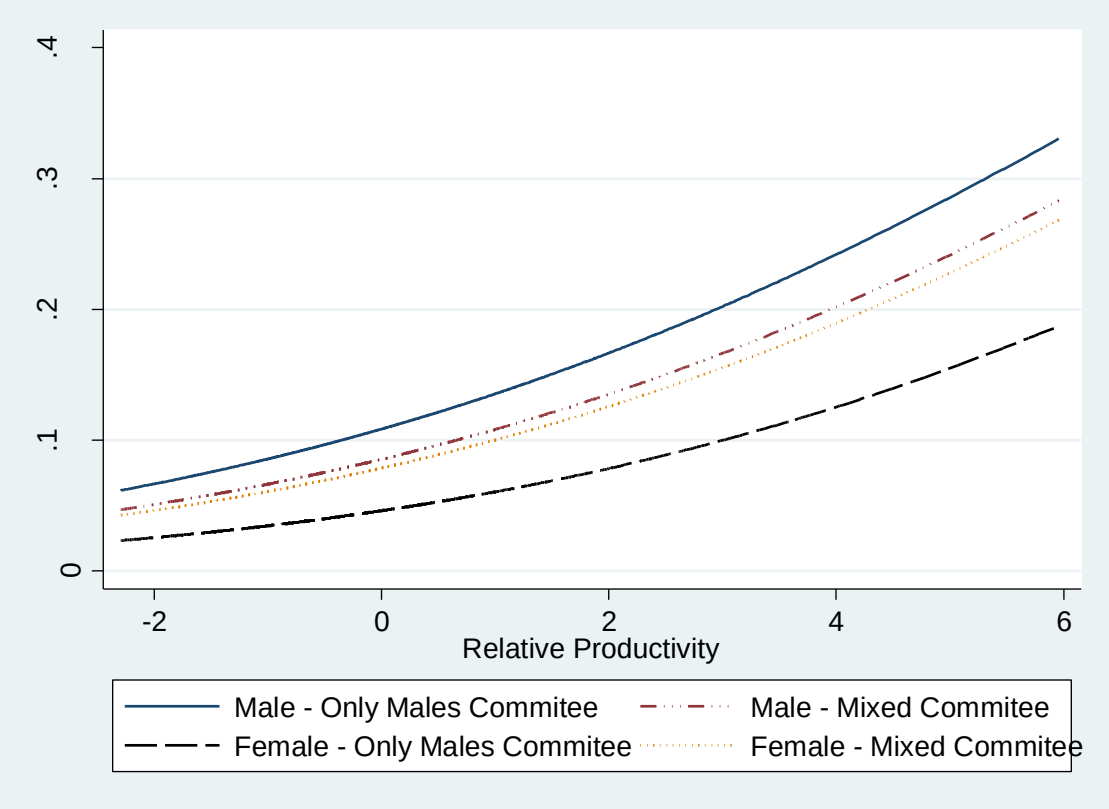

Figure 1. Probability of success of male and female candidates according to the gender composition of the committee

As shown in Table 5 we obtain similar results to those reported in Table 4 when using a linear probability model. The magnitude of the effects and statistical significance are very similar to the marginal effects of Probit estimates in Table 4.

Table 5. Estimates of the Probability of Success. Linear Probability Model.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.050***	-0.045***	-0.079***	-0.073***	-0.074***	-0.060***	-0.072***
	(0.014)	(0.014)	(0.017)	(0.018)	(0.018)	(0.017)	(0.018)
Female*(Females in Com.)			0.063**	0.059**	0.059**		0.059**
			(0.027)	(0.026)	(0.026)		(0.026)
Female*(% Females in Com.)						0.091	
						(0.067)	
Females in Com.			-0.025**	-0.024*	-0.024*		-0.024*
			(0.011)	(0.012)	(0.012)		(0.012)
% Females in Com.		0.024***		0.024***		0.024***	0.025***
		(0.004)		(0.004)		(0.004)	(0.005)
Relative Productivity		0.275***		0.274***	0.274***	0.274***	0.275***

and with it the probability of winning the competition the gender of the candidate produces a relevant effect on promotion chances.

		(0.027)		(0.027)	(0.027)	(0.027)	(0.027)
Insider		0.058***		0.058***	0.057***	0.057***	0.058***
		(0.021)		(0.021)	(0.021)	(0.021)	(0.021)
Connections		0.001		0.002	-0.003	0.002	0.002
		(0.018)		(0.018)	(0.018)	(0.018)	(0.018)
University Job		-0.001		-0.001	-0.001	-0.001	-0.001
		(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
Age					0.009***		
					(0.002)		
Relative h-index						-0.031	
						(0.030)	
(Relative Prod.)* (Females in Com.)							-0.004
							(0.008)
(Relative Prod.)* (Females in Com.)*Female							0.006
							(0.012)
Observations	2279	2279	2279	2279	2279	2279	2279
Adj. R-squared	0.023	0.126	0.024	0.128	0.126	0.126	0.127

Notes: The dependent variable is *Success*. Linear Probability Model. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

We have also experimented using multi-way clustering for standard errors, as suggested by Cameron *et al.* (2006), at competition and candidate level. The significance of the coefficients of interest does not change and, if anything, it slightly improves.

IV. Are the Effects of Mixed Sex Committee Heterogeneous across Fields and Positions?

In this section we investigate whether the effects of a mixed sex committee are heterogeneous across fields and in relation to the type of position for which promotion is decided. We re-estimate specification 4 of Table 4 separately for competitions to associate and to full professor positions (respectively column 1 and column 2 of Table 6). It emerges that in competitions to associate professor committees composed exclusively by males operate a stronger discrimination against women in comparison to that emerging in competitions to full professor positions. Moreover, the improvement produced by a mixed sex committee is smaller in magnitude in competitions to associate professor.

More precisely, in the competitions for associate professor positions, when evaluators are exclusively males, females experiment a reduction in the probability of success of 8 percentage points, while the presence of a mixed sex committee reduces the bias against women to 4.7 percentage points (statistically significant at the 5 percent level). On the other hand, in competitions to full professor, when evaluated by an all-males committee, females' candidates face a reduction in the probability of success of

5 percentage points. In this type of competition, the bias against women vanishes when the evaluation committee is composed also by female members: the difference in the probability of success between males and females is 0.031 with a *p*-value of 0.280.

All in all, these results suggest that in both type of competitions the presence of female evaluators enhances the probability of success of female candidates and helps at reducing the bias against women produced by all-males committees, but the effect is smaller in magnitude for competitions to associate professor positions.

In columns 3 and 4 of Table 6 are reported estimation results for competitions taking place in the fields of, respectively, Economics and Chemistry. All-male committees are gender biased in both fields, but the bias is larger in competitions taking place within the Chemistry field: females experiment a reduction in the probability of success of 7 percentage points in Chemistry and 5 percentage points in Economics. The effect of a mixed sex committee goes in the same direction in both fields and allows male and female candidates to face equality of treatment.

We have also tried to investigate whether results are different depending on the number of females in the field. Preliminarily, we note that the distinction between fields with high and low presence of females almost overlaps with the distinction between Chemistry (42% of females) and Economics (28% of females). However, to better investigate this issue we include directly in the regression the *Percentage of Females in the sub-field* and use the interaction between *Female* (Percentage of Females in the sub-field, demeaned)*. We find that these two variables are far from being statistically significant. On the other hand, the coefficients on our two variables of interest, *Female* and *Female*(Females in Comm.)*, remain very similar to previous findings, suggesting that the effect does not vary according to the number of females in the same field (results not reported).

Table 6. Heterogeneous Effects across Positions and Fields

	(1) Associate Professor	(2) Full Professor	(3) Economics	(4) Chemistry
Female	-0.080***	-0.050**	-0.051**	-0.070***
	(0.024)	(0.021)	(0.021)	(0.023)
Female*(Females in Com.)	0.038	0.120*	0.063	0.070*
	(0.041)	(0.067)	(0.050)	(0.048)
Females in Com.	-0.007	-0.028**	-0.020	-0.020
	(0.015)	(0.014)	(0.013)	(0.015)
Relative Productivity	0.024***	0.015***	0.024***	0.017***
	(0.005)	(0.003)	(0.004)	(0.004)
Insider	0.292***	0.263***	0.222***	0.329***
	(0.038)	(0.051)	(0.046)	(0.037)
Connections	0.065*	0.067*	0.112***	0.049
	(0.034)	(0.038)	(0.042)	(0.034)
University Job	0.018	-0.047	0.011	-0.017
	(0.023)	(0.037)	(0.022)	(0.035)
Age	0.003**	-0.004***	-0.001	-0.001
	(0.001)	(0.001)	(0.002)	(0.001)
Observations	1255	1024	1021	1258
Pseudo R-squared	0.182	0.181	0.154	0.188

Notes: The Table reports marginal effects of Probit estimates (evaluated at the mean values of the explanatory variables in the sample). The dependent variable is *Success*. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

All in all, our evidence is suggestive of some form of taste-based gender discrimination: the positive effect of having a female member in the committee on female candidates' probability of success may depend on the fact that female evaluators are less likely to have a distaste for female candidates. Since we have found quite similar results across different fields, it seems hard to argue that the discrimination of male evaluators against female candidates is related to unobserved quality or to the fact that male candidates perform better in some unobserved "task" that are particularly appreciated by males.

On the other hand, the larger gender gap in competitions for associate professor positions and the smaller effect of female evaluators on female candidates' probability of success points to a role also for statistical discrimination. Presumably, the signal of a candidate's ability embodied by his/her research productivity in competitions to associate professor positions is weaker and evaluators may use gender in order to infer ability. However, these are only suggestive explanations and other explanations are possible: for example, as we only observe candidates applying for a position, we cannot exclude that the difference in discrimination emerging between competitions to associate and full professor positions might be due to the fact that only more productive women apply for a full professor position.[11]

[11] Since all evaluators in both type of competitions are full professors (and they are not any more in competition with other professors), the different effect we find on competitions to associate professor positions cannot be due to the fact that committee members are afraid of the future competition of successful candidates.

The larger gender gap in competitions for associate professor positions could also be related to the fact that the gender of candidates, even if it is known to the evaluators in both types of competitions, becomes more evident in competitions to associate professor (since candidates have to perform in a lecture and interview) and this, in line with results obtained by Goldin and Rouse (2000), could lead to a stronger discrimination.

We have also tried to investigate whether the stronger effect produced by the gender composition of the committee on evaluations to full professor positions is suggestive of the fact that female candidates' probability of success instead of being affect by the gender composition of evaluators is related to being part of networks (which could be affected by gender and by age). Since networks could be stronger for more senior candidates, we have run our main regression (Table 4, specification 4, with the full set of controls) separately for candidates younger and older than the median age in each type of position. We find that the effect of having a mixed evaluation committee on female candidates' probability of success is similar for younger and older candidates (estimates not reported) suggesting that the effect of gender composition on females' probability of success does not derive from being part of the same network.

As our measures of individual productivity (based on the number of publications, citations, *h* and *g* indexes) could not adequately take into account the scientific quality of each candidate, we have alternatively measured their productivity considering the impact factor of the Journals in which they publish. Limitedly to the field of Economics, using the Journal Citation Reports, a product of Thomson ISI (Institute for Scientific Information) for the year 2009, we have attributed to each publication of each candidate the impact factor of the publishing Journal using both the impact factor (for the last two years) and the 5-year impact factor and we have calculated the *Relative Impact Factor* as the difference between a candidate's Impact Factor and competitors' Impact Factor. Also in this case we find that female candidates have a lower probability of success in case of an all-male committee, while the negative effect for females vanishes when there is at least a female among committee members (results not reported).

For the field of Economics, we have also measured productivity using exclusively recent publications (for the period 2004-2010). Considering recent productivity may be relevant if women are more likely to experience a slowdown in productivity as they enter their midcareer phase, since in this phase their family-rearing requirements become typically more pressing. We build a variable *Recent*

Relative Productivity – following the same procedure we used for the variable *Relative Productivity* using the number of papers, citations and impact factor since 2004. We find that *Recent Productivity* is highly correlated to *Productivity*: the rate of correlation is 0.75. In addition, we find that the impact of *Relative Recent Productivity* is very similar to the impact of the comprehensive *Relative Productivity*. The statistical significance is also similar. More importantly, the coefficients on *Female* and on *Female*(Females in Comm.)* remain almost unchanged.

V. Robustness Checks

In this Section we present the results of a number of robustness checks. Firstly, we investigate whether our results are affected by the candidates' withdrawals decisions. In principle, males' and females' expectations might differ ex ante and differences in expectations may affect outcomes ex post. More specifically, female candidates might believe that they are less likely to be promoted when the evaluating committee is composed exclusively by men and, as a consequence, once informed about the gender composition of the committee, they may decide to withdraw from competition. In this case, the outcome we observe would be driven by women's expectations rather than be determined by the effective behavior of all-male committees. Secondly, we check whether our findings are driven by the fact that female and male evaluators have different characteristics in terms of scientific productivity, geographical provenience and age. We also verify whether the results are affected by the characteristics of the internal member and whether they are related to the fact that males and females tend to sort in different research subjects and evaluators tend to be sympathetic toward candidates sharing their own research interests.

Dealing with Withdrawals

Thanks to the availability of information on the behavior of each candidate, we are able to check whether our results are driven by the fact that female candidates retire once they knew the gender composition of the committee. At this aim, we have excluded from our sample all the candidates who have withdrawn from competition (about 27% of candidates).

In Table 7 we present the first four specifications reported in Table 4, plus the results obtained separately for competitions to associate and full professor positions, considering only the sample of

candidates that have maintained their candidacy until the end of the competition process. Our previous findings are confirmed. Again, it emerges that committees composed exclusively by males discriminate against women reducing their probability of success: the effect is even higher in magnitude than that emerging from previous estimates based on the full sample of applicants. On the other hand, mixed sex committees tend to reduce gender discrimination. Therefore, our results are not driven by self-fulfilling expectations.

Results similar to those reported in columns 1 and 2 of Table 6 are found considering separately competitions to associate and full professor: discrimination against women is stronger in competitions to associate professor, where we also observe a smaller positive impact of mixed sex committees compared to that emerging in competitions to full professor. The magnitude of the effects is similar to that found when considering also candidates who have withdrawn from competition, implying that withdrawals play only a minor role in explaining the phenomenon we study.

These findings suggest that differences between males and females in preferences for competition (see Bertrand, 2011) play a minor role in explaining female worse outcomes. In fact, in the estimates of Table 7 we have considered subjects who have applied for a position and who have maintained their candidacy until the conclusion of the evaluation process showing in this way their willingness to compete. However, we are aware that withdrawal decisions are driven by many potential reasons and then we take this evidence as being only suggestive.

Table 7. Estimates of the Probability of Success Excluding Withdrawals

	(1) All	(2) All	(3) All	(4) All	(5) Associate Professor	(6) Full Professor
Female	-0.063***	-0.050***	-0.098***	-0.085***	-0.143***	-0.056***
	(0.018)	(0.017)	(0.022)	(0.022)	(0.047)	(0.022)
Female*(Females in Com.)			0.087**	0.092*	0.063	0.115**
			(0.047)	(0.048)	(0.074)	(0.067)
Females in Com.			-0.027**	-0.025*	-0.006	-0.029**
			(0.014)	(0.014)	(0.029)	(0.014)
Relative Productivity		0.026***		0.026***	0.043***	0.016***
		(0.004)		(0.004)	(0.013)	(0.003)
Insider		0.282***		0.283***	0.304***	0.264***
		(0.034)		(0.034)	(0.050)	(0.052)
Connections		0.085***		0.085***	0.074	0.078***
		(0.033)		(0.034)	(0.054)	(0.041)
University Job		-0.035		-0.034	-0.039	-0.100
		(0.035)		(0.035)	(0.049)	(0.052)
Age		-0.002		-0.002	0.006***	-0.005
		(0.002)		(0.002)	(0.003)	(0.002)
Observations	1652	1652	1652	1656	708	944
Pseudo R-squared	0.079	0.183	0.082	0.187	0.171	0.192

Notes: The Table reports marginal effects of Probit estimates (evaluated at the mean values of the explanatory variables in the sample). The dependent variable is *Success*. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

To better understand candidates' decisions of withdrawing from competition we have estimated a model for the probability of withdrawal. It emerges that females are more likely to withdraw from competition than their male counterparts. The withdrawal decision, however, is not affected by the committee gender competitions (see Appendix B).

Other committee characteristics

As shown in Table 2 committee members show some differences in a number of characteristics. To check whether our results might be affected by these differences, we have interacted the dummy *Female* with measures of evaluators' productivity, age and working geographical area.

In column 1 of Table 8 we re-estimate specification 4 of Table 4 adding as controls a dummy variable, *Highly Productive Committee*, taking the value of 1 for committees with an average quality in terms of scientific productivity (measured with our comprehensive measure *Scientific Productivity*) above the mean and an interaction term between this variable and the dummy *Female*. This allow us to investigate whether female candidates are more or less favored by evaluators with different research quality. We find that the sign of the interaction coefficient is negative, but the effect in far from being statistically significant. Nevertheless, adding these controls does not change our results on the effects of the gender committee composition on discrimination against female candidates (which hold true also when we measure the research quality of both candidates and committee members using the *h* or the *g* index - not reported).

In column 2 of Table 8 we report estimates obtained considering the effect produced by evaluators of different age on female candidates' probability of success. As attitudes toward gender roles may change over time, it could be that older generations are more female averse while younger ones are less likely to discriminate against women. If male evaluators are older compared to their female counterparts, it could be that the positive effect of mixed sex committees on female candidates' probability of success is not related to the gender composition of the committee but to the age of its members. To investigate this aspect we have added among our controls an interaction term between the dummy *Female* and a dummy

Above Mean Age Committee taking the value of 1 when the average age of committee members is above the mean in the sample and zero otherwise. As shown in column 2, the age of evaluators has no effect on the probability of success of female candidates, while the committee gender composition continues to produce the effects discussed above.

We also investigate whether evaluators working in different areas of the country show different attitudes toward women. As shown by a number of studies, women are more likely to be relegated to traditional roles in the South of Italy. As a consequence, we may expect that males working in Southern regions are more likely to be affected by gender stereotypes and to discriminate against women. At the purpose of investigating this issue, we consider the fraction of evaluators working in Southern regions (% of evaluators from the South) and interact this variable with the dummy *Female*. As shown in column 3 of Table 8, female candidates are slightly more likely to suffer discrimination when the evaluators work in universities located in the South of Italy (although the *p*-value is only 0.203). However, no relevant change is observed as regards the effect of the committee gender composition on females' probability of success.

In column 4 we check the robustness of our results controlling for all the committee characteristics described above. Again our main results remain substantially unchanged.

In column 5 we have also included among controls the characteristics of the internal committee member in terms of gender and scientific productivity and have interacted these features with the dummy *Female*. The results of interest remain substantially unchanged.

We have also estimated separately our main specification for competitions in which the internal member is a male and for competitions in which the internal member is a female (results not reported). When the internal member is a male, we find almost identical results to our main findings. On the other hand, when the internal member is a female but all the other commissioners are males, we find a slightly lower coefficient on *Female* (-0.054 instead of -0.064) but the *p*-value is only 0.15. In addition, with a female internal commissioner, the effect of having a mixed committees on female candidates' probability of success becomes stronger (0.14), but again it is not statistically significant at conventional levels (*p*-value=0.21). These rather imprecise estimates are probably due to the fact that we have only 16 competitions with a female internal member.

Table 8. Controlling for other committee characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.056***	-0.044*	-0.063***	0.033	-0.056***	-0.048**
	(0.018)	(0.023)	(0.022)	(0.029)	(0.018)	(0.021)
Female*(Females in Committee)	0.074**	0.064*	0.072**	0.066*	0.074**	0.051
	(0.036)	(0.036)	(0.036)	(0.036)	(0.036)	(0.049)
Females in Committee	-0.021**	-0.019*	-0.020**	-0.020**	-0.021**	-0.023
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.014)
Female*(Highly Productive Committee)	-0.021			-0.023	-0.021	
	(0.023)			(0.022)	(0.023)	
Highly Productive Committee	0.014			0.014	0.014	
	(0.010)			(0.010)	(0.010)	
Female*(Above Mean Age Committee)		-0.001		-0.005		
		(0.026)		(0.025)		
Above Mean Age Committee		-0.006		-0.004		
		(0.011)		(0.011)		
Female*(% of Evaluators from the South)			-0.065	-0.066		
			(0.051)	(0.049)		
% of Evaluators from the South			0.015	0.013		
			(0.020)	(0.020)		
Productivity Internal Member					0.001	
					(0.002)	
Female Internal Member					0.001	
					(0.007)	
Female*(Female Internal Member)					0.035	
					(0.039)	
Female*(Productivity Internal Member)					-0.007	
					(0.005)	
Same Subject						0.031
						(0.044)
Female*(Same Subject)						-0.012
						(0.070)
Female*(Same Subject)* (Females in Committee)						0.027
						(0.066)
(Same Subject)* (Females in Committee)						0.026
						(0.129)
Observations	2279	2279	2279	2279	2279	1021
Pseudo R-squared	0.171	0.171	0.170	0.172	0.171	0.161

Notes: The Table reports marginal effects of Probit estimates (evaluated at the mean values of the explanatory variables in the sample). The dependent variable is *Success*. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.

As evaluators could be sympathetic to candidates working in their same research subject (see Hamermesh and Schmidt 2003, Bagues and Villadoniga 2009), if males and females tend to specialize in different subjects of research (as shown by Dolado et al. 2012), our finding that evaluators tend to favor candidates of their own gender could be due to a common subject effect.

To investigate this aspect we focus our attention on Economics, which we know better and for which we are able to classify journals according to the research subjects. To analyze the effect produced by the candidate sharing the same research interest of the evaluators on his probability of winning the competition, we have firstly classified in 22 different subject categories (such as Finance; Macroeconomics; Public Economics; Education and Labor; Business Economics; International Economics

and so on) all the economics Journals with an Impact Factor (from the Journal Citation Report, year 2009). Secondly, we have attached a subject of research to each commissioner and to each candidate on the basis of their greatest number of publications within a subject category. Thirdly, we have defined a dummy *Same Subject* which takes the value of one if in the evaluation committee there is at least one member who publishes on the same subject of the candidate.

We estimate the specification in column 4, Table 4, adding as explanatory variables *Same Subject* and two interaction terms *Female*Same Subject* and *Female*Same Subject*Females in Committee*. Estimation results are reported in Table 8, column 6. *Same Subject* is positive (although not statistically significant at the standard levels), suggesting the existence of a "similar-to-me" effect: a candidate has a higher probability of success if he/she is assigned to a committee in which a member works in his/her same subject of research. Given that the interaction term *Female*Same Subject* is not significant, the "same subject" effect is almost the same for males and females. The interaction term *Female*Same Subject*Females in Committee* is far from being statistically significant. Our main findings remain substantially unchanged also when we control for these additional variables: females are disfavored if they are evaluated by an all-male committee, while the probability of success of males and females does not differ significantly if in the committee there is at least a female evaluator (this both for females working in the same field of their evaluators than for female candidates working in a different field).[12]

VI. Concluding Remarks

Females typically obtain worse results compared to their male counterparts in many dimensions of social and economic life. A large empirical evidence shows that females earn substantially less than males even when they perform the same job and have the same qualification. In addition, the presence of females in top and high-ranking positions is negligible in many countries.

[12] As according to the rules followed for the composition of evaluation committees, in sub-fields where the number of opened vacancies was small compared to the number of available evaluators, committee members had to be randomly selected among a number of professors elected by the professors in the sub-field, we have checked the robustness of our findings considering exclusively those competitions in which the evaluators were randomly selected from the whole body of full professors in the sub-field. Results are consistent with those found considering the whole sample of competitions.

This state of affairs explains why policymakers and researchers often debate about what types of policies may promote gender equality. In the past, equal opportunities and equal treatment laws have been the main focus, reflecting the widespread idea that improvement in women's access to education would have allowed them to reach positions similar to those held by men. However, this kind of policy has not produced the expected results and unbalances in top and influential positions still persist. Recently, in order to overcome these inequalities many countries have introduced gender parity in top positions. Whether the hiring or promotion of more women to influential positions represents an effective way to break the "glass ceiling" for females is still a matter of discussion (Chattopadhyay and Duflo 2004; Pande 2003; De Paola et al. 2010).

In this paper we have tried to shed some light on this issue focusing on female performance in academic promotions and trying to understand whether the gender of evaluators matters. Relying on a large randomized natural experiment consisting in the examinations for promotion to associate and full professor positions in the Italian University, where the allocation of evaluators to each competition was random, we have investigated the candidates' probability of success and how it is affected by the gender composition of evaluation committees.

From our analysis it emerges that, even after controlling for individual characteristics, measures of scientific productivity and indicators of social connections, females experiment a considerable lower probability of success. Interestingly, females' chances of success are affected by the gender of evaluators. In competitions in which the evaluators are exclusively males, female candidates suffer a reduction of their probability of success of about 6 percentage points: this implies that the probability of success of females is about 50% lower than males. On the other hand, gender discrimination almost vanishes when the candidates are judged by a mixed sex committee. We find very similar results across different types of positions and different fields.

In addition, the discrimination against females operated by all-male committees and the positive impact of mixed sex committees on female candidates' probability of success persists also when we exclude from our sample the candidates who have withdrawn from competition. Females might believe that they are less likely to succeed when the evaluation committee is composed exclusively by men and decide to retire from competition when facing an all-male committee. Nevertheless, we do not find

evidence that the committee gender composition affects the candidates' probability of withdrawing from competition.

Our findings showing that a greater number of females in influential positions may help other women to advance in their career are in line with those emerging from Zinovyeva and Bagues (2011) showing, at least for competitions to full professor in Spain, a positive effect of female evaluators on the probability of success of female candidates, but are in contrast with the findings of Bagues and Esteve-Volart (2010) who, for the access to positions in the Spanish Judiciary, find that women in evaluation committees favor male candidates.

These diverging results suggest that the attitudes of mixed sex committees toward male and female candidates may change in relation to different contexts or in relation to the positions to be filled. Additional research seems necessary in order to better understand the role of females in preventing gender discrimination and to formulate policy recommendations aimed at promoting equality of treatment for men and women.

References

AIGNER, D. and CAIN, G. (1977). Statistical Theories of Discrimination in Labor Markets. *Industrial and Labor Relations Review*, **30**, 175-187.

ALBRECHT, J., BJORKLUND, A., and VROMAN, S. (2003) Is there a glass ceiling in Sweden? *Journal of Labor Economics*, **21**, 145–177.

ALTONJI, J. and BLANK, R., (1999). Race and gender in the labor market. In O.Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*. Volume 3C, 3143–259, Amsterdam: North-Holland.

ARULAMPALAM, W., BOOTH, A., and BRYAN, L. (2007). Is there a glass ceiling over Europe? Exploring the gender pay gap across the wages distribution. *Industrial and Labor Relations Review*, **60**, 163–186.

BAGUES, M and ESTEVE-VOLART, B. (2010). Can Gender Parity Break the Glass Ceiling? Evidence from a Repeated Randomized Experiment. *Review of Economic Studies*, **77**, 1301-1328.

____ and PEREZ-VILLADONIGA, M. (2009). Do recruiters prefer applicants with similar skills? Evidence from a randomized natural experiment. *Business Economics Working Papers*, Universidad Carlos III, Departamento de Economía de la Empresa.

BECKER, G. (1957). *The Economics of Discrimination*, Chicago: University of Chicago Press.

BELL, L. (2005). Women-Led Firms and the Gender Gap in Top Executive Jobs. *IZA Discussion Paper* 1689.

BERTRAND, M. (2011). New Perspectives on Gender. *Handbook of Labor Economics*, vol. 4b.

BLACKABY, D., BOOTH, A., and FRANK, J. (2005). Outside offers and the gender pay gap. *Economic Journal*, **115**, F81–F107.

BLAU, F., and KAHN, M. (2003). Understanding international differences in the gender pay gap. *Journal of Labor Economics*, **21**, 106–144.

BOOTH, A. (2009). Gender and Competition. *Labour Economics*, **16**, 599–606.

CAHUC, P. and ZYLBERBERG, A. (2004). *Labor Economics*. MIT Press.

CAMERON, C., GELBACH, J. and MILLER, D. (2006). Robust Inference with Multi-way Clustering. *NBER Technical Working Papers* 0327.

CARDOSO, A.R., and WINTER-EBMER, R. (2007). Mentoring and segregation: female-led firms and gender wage policies, *IZA Discussion Paper* No. 3210, December.

CHATTOPADHYAY, R. and DUFLO E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, **72(5)**, 1409–1443.

CHECCHI, D. (1999). Tenure. An Appraisal of a National Selection Process for Associate Professorship. *Giornale degli Economisti*, **58**, 137-181.

COMBES, P, LINNEMER L. and VISSER M. (2008). Publish or peer-rich? The role of skills and networks in hiring economics professors. *Labour Economics*, **15**, 423–441,

DE PAOLA, M., SCOPPA, V., and LOMBARDO, R. (2010). Can Gender Quotas Break Down Negative Stereotypes? Evidence from Changes in Electoral Rules. *Journal of Public Economics*, **94**, 344-353.

DOLADO, J., FELGUEROSO F. and ALMUNIA, M. (2012). Are men and women-economists evenly distributed across research fields? Some new empirical evidence. SERIEs 3.3 (2012): 367-393.

GINTHER, D. and HAYES, K. (2003). Gender Differences in Salary and Promotion for Faculty in the Humanities, 1977-1995. *Journal of Human Resources*, **38 (1)**, 34-73.

___ and KAHN, S. (2004). Women in Economics: Moving Up or Falling Off the Academic Ladder. *Journal of Economic Perspectives*, **18(3)**, 193-214.

___ and KAHN, S. (2009). Does Science Promote Women? Evidence from Academia 1973-2001. NBER chapters in: *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*. 163-194.

GOLDIN, C., and ROUSE, C. (2000). Orchestrating impartiality: the impact of blind auditions on female musicians. *American Economic Review*, **90(4)**, 715–741.

HAMERMESH, D. and SCHMIDT, P. (2003). The Determinants of Econometric Society Fellows Elections. *Econometrica*, **71**, 399-407.

Higher Education Statistics Agency, 2010 (http://www.hesa.ac.uk/)

HILMER, C., and HILMER, M. (2010). Are There Gender Differences in the Job Mobility Patterns of Academic Economists? *American Economic Review*. **100**, 353-357.

LAVY, V. (2008). Do Gender Stereotypes Reduce Girls' Human Capital Outcomes? Evidence from a Natural Experiment. *Journal of Public Economics*. **92(10-11)**, 2083-2105.

MANNING, A. and SAIDI, F. (2010). Understanding the Gender Pay Gap: What's Competition Got to Do with It? *Industrial & Labor Relations Review*, 63(7),

McDOWELl, J. M., SINGELL L. and ZILIAK J. P. (1999). Cracks in the Glass Ceiling: Gender and Promotion in the Economics Profession. *American Economic Review, Papers and Proceedings*, 89(2), 392-396.

PANDE, R. (2003). Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? *American Economic Review*, **93(4)**, 1132-1151.

PEROTTI, R. (2002). The Italian University System: Rules vs. Incentives. Presented at the first conference on "Monitoring Italy". ISAE, Rome.

WEICHSELBAUMER, D. and WINTER-EBMER, R. (2005). A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys*, **19**, 479-511.

WENNERAS, C. and WOLD, A. (1997). Nepotism and sexism in peer-review. *Nature*, **387**, 341-343.

ZINOVYEVA, N. and BAGUES, M. (2011), Does gender matter for academic promotion? Evidence from a randomized natural experiment. *IZA Discussion Paper* 5537.

Appendix A. Females' Promotion Chances before and after the Reform

The sheer change in the system governing promotions in Italian Universities could have affected women's promotion chances. The passage from a system (before 2008) in which the selection of committee members was mainly determined by influential professors (even if committee members were formally elected) to a new one (starting since 2009) with randomly selected committee members may have both improved or hurt women's promotion prospects. On the one hand, randomly selected committee members may be less familiar with the work of the candidates and then may use their gender to infer productivity. On the other hand, the new system has increased the probability of having females among committee members: under the old system elected members were *de facto* chosen among powerful professors who were typically men.

To provide some evidence on this issue we have gathered data on the academic positions held by all Italian professors respectively in 2001, in 2008 and in 2012 (Ministry of Education, University and Research (MIUR) website) totaling to about 60,000 observations per each year. We observe the position at each data for each professor (Assistant Professor, Associate Professor, Full Professor). We build a dummy *Promotion* equal to one if a professor is promoted from assistant professor to associate professor or from associate professor to full professor, respectively, in the period 2001-2008 (before the reform) and 2009-2012 (after the reform). The dummy is set equal to zero if a professor maintains the same position, dealing separately with each of the two periods considered.[13] Those who were full professors in 2001 are excluded.

We then estimate the probability of being promoted for males and females, before and after the reform. Results are reported in Table A1. We use a linear probability model, since the model is fully saturated, controlling for 28 field dummies.

As shown in column (1), before the reform females had a probability of being promoted of 7 percentage points lower than males (*t*-stat=-12.5). After the reform (in column 2) the probability of being promoted for females is only 2.1 lower than males (*t*-stat=-7.3) improving of about 4.9 percentage points. Then, we pool together observations for the two periods and in column 3 we estimate the probability of promotion for males and females before and after the reform. The difference between males and females of 6.5 p.p. before the reform remains significant after (2.4 p.p., *t*-stat=-8.3), but the magnitude is considerably reduced of about 4.1 p.p.: the interaction term *Female*(Post Reform)* is positive and highly statistically significant. Very similar results are obtained if we focus only on the two fields (Economics and Chemistry) analyzed in the paper (column 4).

[13] In this way we are not considering the candidates who were working outside the university system and have applied for a position as associate or full professor. However, these cases are quite rare. In Table 1 of the paper we have shown that about 90% of applicants for an associate or full professor position were performing an academic job and external successful candidates were only a few.

This evidence is only suggestive, both because we are not controlling for candidates' productivity and because of the possible influence of other temporal trends. However, if male and female productivity patterns have not changed in coincidence with the reform and no other relevant changes occurred at that time, our results show that the introduction of randomly selected committees has improved women's promotion chances, probably by reducing the power of those professors (who were typically males) who under the old system had a role in shaping the selection of evaluation committees.

Table A1. Female probability of promotion before and after the reform. Linear Probability Estimates

	(1) Before Reform	(2) After Reform	(3) Pooled Model	(4) Economics and Chemistry
Female	-0.070***	-0.021***	-0.065***	-0.091***
	(0.006)	(0.003)	(0.005)	(0.017)
Female*(Post Reform)			0.041***	0.069***
			(0.006)	(0.019)
Post Reform			-0.303***	-0.359***
			(0.004)	(0.013)
Observations	32355	38138	70493	7698
Adjusted R-squared	0.039	0.008	0.138	0.172

Notes: The Table reports LPM estimates. The dependent variable is *Promotion*. In regressions (1)-(3) we control for 28 field dummies. Standard errors (corrected for heteroskedasticity) are reported in parentheses. The symbols *** indicates that coefficients are statistically significant at the 1 percent level.

Appendix B. Withdrawal Decisions by Candidates

To better understand candidates' decisions of withdrawing from competition we have also estimated a probit model considering as dependent variable a dummy taking value of 1 for candidates deciding to withdraw from competition and zero otherwise. Withdrawals are more frequent in competitions to associate professor positions since participation costs are higher due to the fact that candidates are evaluated not only in relation to their CVs but also considering their performance in a teaching lecture, typically given in the place where is located the university posting the vacancy. To take into account this aspect, we have added to the controls used in previous estimates the dummy variable *Distance* taking the value of one when the university in which the candidate is currently employed is located in a geographical area that is different from that of the university initiating the competition. We exclude for each competition the candidates that have been already promoted in some concluded competition.

In Table B1 are reported estimation results. In column 1 we estimate the difference in the probability of withdrawal between males and females controlling for scientific sub-field dummies, type of position, number of competing candidates and *Distance*, without taking into account other candidates' characteristics. It emerges that females are more likely to withdraw from competition than their male counterparts (+3.4 percentage points). The same result holds true when we add among controls *Relative Productivity*, *Insider* and *Connections* (column 2).

In column 3 we investigate whether the probability of withdrawal is affected by the committee's gender composition. We do not find any statistically significant effect for competitions to associate professors, while we find a negative effect for competitions to full professor.

In columns 4 and 5 we run separate regressions respectively for competitions to associate and to full professor positions. It emerges that females are more likely to withdraw from competitions to associate professor, while it does not emerge any statistically significant difference between males' and females' withdrawal decisions as regards competitions to full professor positions. Moreover, while the gender composition of the committee does not affect withdrawal decisions in competitions to associate professors, we find that in competitions to full professor positions females are less likely to retire their candidacy when the evaluation committee is composed also by female members.

Table B1. The Determinants of Withdrawal from Competition

	(1) All	(2) All	(3) All	(4) Associate Professor	(5) Full Professor
Female	0.034*	0.036*	0.044	0.091**	0.003
	(0.019)	(0.019)	(0.027)	(0.046)	(0.014)
Female*(Females in Com.)			-0.016	-0.012	-0.037*
			(0.038)	(0.065)	(0.012)
Females in Com.			-0.013	0.013	-0.010
			(0.033)	(0.053)	(0.018)
Relative Productivity		-0.009	-0.009	-0.036***	-0.004
		(0.007)	(0.007)	(0.012)	(0.015)
Distance	0.128***	0.049**	0.050**	0.105**	0.045
	(0.019)	(0.023)	(0.023)	(0.041)	(0.031)
Associate Professor	0.327***	0.416***	0.413***		
	(0.026)	(0.034)	(0.036)		
Insider		-0.131***	-0.131***	-0.247***	-0.018
		(0.023)	(0.023)	(0.043)	(0.016)
Connections		-0.068***	-0.068***	-0.131***	-0.012
		(0.022)	(0.022)	(0.040)	(0.018)
University Job		-0.120***	-0.126***	-0.073*	-0.223***
		(0.044)	(0.043)	(0.044)	(0.076)
Age		0.000	-0.000	-0.000	-0.001
		(0.002)	(0.002)	(0.002)	(0.001)
Observations	2090	2090	2090	1146	925
Pseudo R-squared	0.206	0.227	0.229	0.113	0.251

Notes: The Table reports marginal effects of Probit estimates (evaluated at the mean values of the explanatory variables in the sample). The dependent variable is *Withdrawal*. In all regressions we control for sub-field dummies, type of position dummy and the number of candidates. Standard errors (corrected for heteroskedasticity and robust to clusters at the competition level) are reported in parentheses. The symbols ***, **, * indicate that coefficients are statistically significant, respectively, at the 1, 5, and 10 percent level.