SciencesPo
LABORATOIRE INTERDISCIPLINAIRE
D'ÉVALUATION DES POLITIQUES PUBLIQUES

LIEPP Working Paper

Novembre 2018, n°82

Gender Quotas in Hiring Committees: a Boon or a Bane for Women?

Pierre DESCHAMPS
Département d'économie de Sciences Po, LIEPP
Pierre.deschamps@sciencespo.fr

Comment citer cette publication:
Pierre Deschamps, Gender Quotas in Hiring Committees: a Boon or a Bane for Women?, *Sciences Po LIEPP Working Paper* n°82, 2018-11-29.

Gender Quotas in Hiring Committees: a Boon or a Bane for Women?*

Pierre Deschamps†
JOB MARKET PAPER

November 29, 2018

Abstract

Following in the footsteps of similar initiatives at the boardroom level in Norway and other European countries, the French government decided to impose a gender quota in academic hiring committees in 2015. The goal of this paper is to evaluate how this reform changed the way women are ranked by these committees. The reform affected academic disciplines heterogeneously. I contrast the effect of the reform between fields that were significantly affected, and those that already respected the quota before the reform. Drawing on a unique dataset made up of administrative data provided by French universities, I show that the reform significantly worsened both the probability of being hired and the ranks of women, with a treatment effect equivalent to a 4 standard deviation drop in h-index. There is evidence that this is driven mainly by the reaction of men to the reform, since the negative effect of the reform is concentrated in committees that are helmed by men.

(JEL J16, J71)

1 Introduction

Even though women make up the majority of Ph.D candidates in many academic disciplines, they are under-represented in faculty positions, and the gender gap increases as the air gets more rarefied. In the European Union, women made up between 40-60% of PhD graduates, 40% of Assistant Professors, and only 21% of full professors, according to the European Commission's 2015 SHE figures.[1]

*I am very grateful to Etienne Wasmer for his advice and help in starting this project. I also thank Ghazala Azmat, Manuel Bagues, Alain Chenu, Etienne Fize, Daniel Hamermesh, Emeric Henry, Nicolas Jacquemet, Morgane Laouenan, Thierry Mayer, Roland Rathelot and Jose de Sousa for their constructive comments, as well as seminar participants at the SOLE, AFSE, EEA and EALE. I thank Michel Gardette, Isabelle Delacroix, Magali Steinmann, Marion Paoletti, Arnaud Dudek, Bruno Oliveira and Anne Kupiec for their help in accessing the data.

†Pierre Deschamps: Sciences Po, Paris, Département d'Economie, and LIEPP pierre.deschamps@sciencespo.fr.

[1]Chenu and Martin (2016) study the career trajectories of sociology professors in France, our country of interest, and show that only 11% of freshly-minted women sociology assistant professors from 1996-1999 had become full professors by 2012, compared to 35% of men. In 2011, women made up 50% of sociology assistant professors, but only 27% of full professors.

If we believe that discrimination is holding back the careers of women, and that discrimination is caused by men with own-gender preferences, then one way to improve the prospects of women could be to increase their number among the academics making hiring and promotion decisions. A new reform by the French government took this line of reasoning and imposed gender balanced academic recruitment committees.[2] Since 2015, French academic hiring committees in the public sector have to be made up of at least 40% of members from each gender. The aim of this article is to directly test the effect of the reform on the hiring and ranking of women by these committees, and *in fine* whether own-gender preferences exist.

This reform also improves the representation of women,[3] which could have positive effects such as encouraging more women to apply for professorial positions. Nevertheless, gender quotas also come with significant costs. The relatively small number of women eligible to participate in committees will mean that the administrative work of each will substantially increase, potentially harming their ability to publish. As argued by Vernos (2013), when discussing gender quotas in ERC evaluation panels, "quotas might make matters worse by overworking already-stretched female scientists". If gender quotas are not efficient in countering own-gender preferences, then the 2015 reform may end up doing more harm than good.

In what follows, I use administrative data provided by French universities to investigate whether academic recruitment committees that are gender balanced lead to better outcomes for women. Surprisingly, the raw correlations directly contradict the presence of own-gender preferences: the higher the share of women evaluators in a committee, the lower women are ranked, even when controlling for publications and academic connections.

I then directly evaluate the impact of the reform, by exploiting the fact that disciplines are differentially affected by the quota; the quota is more likely to bind in fields and universities where the share of women jurors was lower than 40% before the reform. I compare the ranks of women before and after the reform, using the disciplines and universities that were already respecting the quota as a control group. Women receive worse ranks after the reform, and the effect is significant and very large (equivalent to a

[2] Link to the decree.

[3] See Azmat (2014) for positive and negative aspects of increasing gender representation in decision-making teams.

4 standard deviation decrease in a candidate's h-index). Women are also less likely to be hired in the treated group after the reform. However, since we do not observe the votes of individual jury members, we do not know whether it is women who have opposite-gender preferences, or men who decide to vote against women whenever more women are their co-jurors, and we have to be cautious in interpreting these results. In fact, one element leans towards this second interpretation: the negative effect of the reform in disciplines and universities where women are under-represented is driven almost entirely by committees helmed by men.

There is no effect of the reform on the gender composition of the applicant pool. It seems that women were not encouraged to stay in academia and apply for professorial positions as a result of the reform, at least in the short-term. Another result is that most of the new female members on the committees are internal members. This implies that women have to do more administrative work for their department following the reform. There are no significant effects however on the "quality" of jury members, as proxied by the average h-index of committees, whether for male or female members or the jury as a whole.

These results cast doubt on whether gender quotas in recruiting are effective in solving vertical segregation. It is important to remember that there are many different explanations for vertical segregation in academia, besides discrimination: for instance, less investment in human capital, perhaps due to the unequal distribution of household labour and psychological differences, such as attitudes towards risk and competition.[4] Teasing out which of these potential causes is the most salient is important since misidentifying the true causes of gender disparities can lead to detrimental policies that may actually worsen gender gaps.[5]

This article contributes to the literature that considers the effect of gender quotas on women. Economists have mostly focused so far on analysing the effect of quotas in boardrooms or in the political arena. For instance, Ahern and Dittmar (2012) find that the 2003 reform in Norway, which imposed gender quotas in the boardroom of companies,

[4] See Bertrand (2011) for a survey on the effect of psychological differences on labour market outcomes.

[5] For instance, Antecol, Bedard, and Stearns (2018), show that a well-meaning policy aimed at encouraging women to take up parental leave, a gender-neutral tenure clock stop, reduced the rate of women receiving tenure by 22% and increased the rate for men by 19%.

had a negative impact on firm value. Bertrand, Black, Jensen, and Lleras-Muney (2018) analyse the effect of the same reform on the labour market outcomes of women. They find that the reform has a positive effect on the wages of women in the boardroom, and that these women are more qualified than before, but that these gains do not trickle-down the wage ladder.[6] My article shows that quotas in academic hiring committees do not have a positive effect on the hiring of women, similar to the results of Bertrand *et al* (2018) in another labour market context.

More generally, this article looks at the effect of the feminisation of evaluators on the prospects of women. Some articles try to study this effect indirectly. For instance, a recent study in France by Breda and Hillion (2016) showed that women were favoured in oral exams for posts as secondary-school teachers in disciplines that had few women instructors. The articles that have tried to study a direct effect have found conflicting results. Some find that having women evaluators has a positive effect on the outcomes for women, on field data (Lincoln, Pincus, Koster, and Leboy (2012), Boring (2017), Zeltzer (2015))[7] or audit studies (Edo, Jacquemet, and Yannelis (2015)).[8] A few find no effect (Abrevaya and Hamermesh (2012), Feld, Salamanca, and Hamermesh (2016), Williams and Ceci (2015)).[9] Ellemers *et al.* (2004), and Broder (1993), find opposite-gender preferences, in the evaluation of doctoral students' work ethics and grant applications (using evaluations from men and women on the same paper) respectively. Whether own- or opposite-gender preferences exist in practice remains an open question.

[6]From articles that look at quotas in politics, Casas-Arce and Saiz (2015) find that following a reform that imposed quotas on electoral lists in Spain, parties that previously had few women candidates had a very large (6 percentage points) increase in vote shares. Beaman *et al* (2009) and (2012) show that in villages which had been randomly selected to have mandatory leadership positions for women in village councils, women were more likely to win local elections in the years that followed, and that voter attitudes towards women in leadership positions improved, as well as the aspirations and educational achievements of women in these villages.

[7]The first article finds that having a male committee chair for research prizes raises the probability of men receiving a prize. Boring (2017) shows that male students give higher evaluation grades to their male professors. Zeltzer finds that male physicians refer patients to male specialists relatively more often than do female doctors.

[8]They find that female recruiters are more likely to call back women.

[9]Abrevaya and Hamermesh (2012) look at whether female referees accept papers written by women more often, using article and reviewer fixed effects, and find no effect. Feld, Salamanca, and Hamermesh (2016) compare students that are graded anonymously with students that are not, which enables them to distinguish between endophily and exophobia. They find evidence of endophily if instructors and students have the same nationality, but no effect for gender. Williams and Ceci (2015) run lab experiments, and find that women and men in biology, engineering and psychology (not in economics however) prefer to hire women over men with identical profiles. These preferences do not change depending on whether the evaluator is male or female.

Similarly, studies evaluating the effect of the gender of candidates and recruiters on the ranking of applicants find contradicting results. For instance Bagues and Esteve-Volart (2010), using data from law recruitments in Spain, find that having more women in committees does not increase the probability of women being hired, whereas Paola and Scoppa (2015), looking at recruitments of Economics and Biology professors, find that having at least one woman in the committee increases the probability of women being promoted. Bagues, Sylos-Labini, and Zinovyeva (2017) use data on 100 000 applications to the position of professor in Italy and Spain. They find that having more women in the recruiting committee harms the chances of women being promoted, except for full professorships in Spain.

My work builds on Bagues *et al* (2017) in several ways. In their data, candidates do not know their jurors, and are unlikely to work with them after the committee has convened. In the committees I study, half of the jurors are professors in the university and will have the highest-ranked candidate as a colleague, which could raise the stakes of hiring, when compared to the decision of promoting an assistant professor to full professor in another university. Moreover, I use ranked data instead of data on promotion decisions, which allows for identification of effects over the whole range of candidates, instead of the marginal candidates who are the only ones harmed by discrimination in the context of promotions. A candidate who is always promoted or always rejected can also be affected by gender bias, but this effect would not be identified in promotion decisions, and relying on estimates from these studies could lead us to underestimate the size of gender preferences. Studying ranks on the other hand, allows us to estimate discrimination parameters whatever the candidate's ability or likelihood of being hired.

One problem in identifying preferences is possible endogeneity in the assignment of candidates to a jury. For instance, candidates could decide not to apply for a position depending on the gender composition of the jury. Bagues *et al* (2017), Bagues and Esteve-Volart (2010), and Paola and Scoppa (2015) all use random assignment of candidates to committees to take care of this concern. I rely instead on variation across time and fields from a policy reform. The difference between the two approaches is relevant if we believe that random assignment in itself may increase or decrease the probability of observing biases,[10] especially since in most labour market contexts, evaluations do not rely on random assignments of jurors to candidates. Furthermore, beyond the identification of own-or opposite-gender preferences, my article gives evidence on several dimensions of the reform (reaction of the gender composition of applicant pools to the quota, and increasing administrative work for internal members) upon which the previous papers cannot contribute. This article can be seen as complementing previous articles that relied on random assignment of jurors to applicants for identification.

In section 2, I explain how I construct the data set, and present some descriptive statistics. In section 3, I discuss how own- or opposite-gender preferences could intervene in the ranking of committees, and how we can estimate them. In section 4, I present results on the hiring and ranks of candidates. In section 5, I look at how the quota affected hiring, and discuss the interpretation of these results in section 6. I conclude in section 7.

2 Data

In this section, I present the dataset compiled from administrative data on 455 hiring committees from 3 different French universities.[11]

2.1 French hiring committees

French academic hiring committees are created *ad hoc* for each position that has to be filled. These committees have a jury president, who has broad powers over the nomination of committee members. The president also has a deciding vote in case of a tie between candidates. There is some variation in the number of members in these committees; by law there must be between 8 and 20 jurors. At least half of the members must be from outside the hiring university. Once a committee has been created, candidates can apply *via* a web platform called GALAXIE,[12] and post their CVs. In some cases (but not all),

[10]See Li (2017), for evidence of a bias/efficiency trade-off when considering the assignment of candidates to jurors in scientific funding.

[11]The data spans from 2009 to 2018, though not necessarily for each of the universities.

[12]Link to GALAXIE and available positions.

they are aware of the jury composition at the time of their application.[13]

As part of the French Government's push for greater gender equality, gender quotas have been introduced in these hiring committees. Since the 1st January 2015, each committee must also be made up of at least 40%[14] of each gender. However, there are neither constraints on the gender of the jury president, nor quotas on the gender of the candidates hired.

The committee then decides which candidates can be auditioned. Once these candidates are auditioned, the committee then makes a decision and ranks the candidates it deems worthy (if any) of being hired by the committee. The committee votes on the final ranking, and not on the candidates individually. Candidates can dissent during the final vote, but in practice, even if committee members disagree during the deliberation process, the final ranking is almost always accepted unanimously. This means that we cannot recover the behaviour of individual jury members, so I can only infer the behaviour of women in committees from two variables: the gender of the jury president, and the percentage of women in the hiring committee.

After the ranking decision is made, if the first-ranked candidate refuses an offer, the position is then offered to the second-ranked candidate, then to the third-ranked candidate if the second also refuses, and so forth. All ranked candidates can potentially receive an offer from the university, if the candidates ranked above them refuse the offer.[15] Therefore, if the committee decides that these candidates do not meet the requirements needed to work at the university some of the auditioned candidates are not ranked. The candidates are aware of the ranks that they receive, which helps them to inform their decision on which offers they should accept or reject. These ranks may also give a signal to candidates on whether it is relevant to pursue an academic career. This is important in our context since Geuna and Shibayama (2015) show that women are more likely to drop out of academia than men.

[13] We could be worried by auto-censure mechanisms, although if these mechanisms exist, they would reinforce our results. Paola and Scoppa (2015) find that women tend to drop out when applying for positions more often than men, but this effect is independent of the composition of the committee.

[14] For maths, this ratio was dropped exceptionally to 14% and to 33% for Political Science until 2016 (Link to the decree). In other fields, this ratio is binding but once the committee has been approved, if a committee member drops out, the replacements do not need to respect the gender quota. The data includes the final composition of the committee.

[15] This is relative infrequent in the data set. Only 20 first-ranked candidates refuse the position.

Table 1: Descriptive statistics on recruitment

	# recruitments	% of women recruited	% of female presidents	% of women in the committee
Biology	65	0.51	0.38	0.47
Business	18	0.33	0.22	0.49
Chemistry	18	0.39	0.44	0.44
Economics	27	0.19	0.08	0.41
Education	18	0.50	0.17	0.46
Engineering	18	0.06	0.11	0.40
History	12	0.33	0.42	0.36
Languages	40	0.47	0.47	0.52
Law	33	0.42	0.12	0.43
Maths	57	0.14	0.30	0.38
Pharmacology	30	0.57	0.40	0.49
Physical Education	13	0.15	0.08	0.46
Physics	28	0.21	0.14	0.40
Political Science	19	0.26	0.32	0.37
Psychology	30	0.57	0.43	0.50
Sociology	29	0.45	0.40	0.37
Total	455	0.36	0.30	0.44

The statistics above are compiled on a dataset that includes assistant and full professors, over the years 2009-2018.

For each committee, I have access to the names and ages of the jurors and the names and ranks of the candidates, as well as the gender composition of the candidate pool in 85% of the cases. During the analysis, the ranking variable includes the K-ranked candidates, and the candidates that audition but are not ranked, who are all given the rank K+1. Since I only use order and not distance in my regressions, this normalisation is harmless. Candidates that are not auditioned are discarded for the regression analysis, but are included when computing the gender ratio of applicants.

2.2 Descriptive Statistics

I present some statistics on the committees in Table 1. As we can see, the number of female presidents and the share of women in committees varies strongly between fields. In some fields, female presidents are few and far between: Only 4 out of 28 recruitments in Physics for instance, or 2 in economics. In general, the percentage of women hired is correlated with the percentage of female presidents and the share of women in committees at the field level, but also with the percentage of female applicants.

For a small (72) number of contests, I am missing data on applicants, and use instead the percentage of women among the ranked candidates. The results from the main specifications are quantitatively similar when these observations are excluded. As shown in Table 2, which excludes these committees, the average share of female applicants is also a good predictor of average percentage of women ranked, especially when we consider only the fields for which we can observe more than 20 committees. An important takeaway from these two tables is that these raw figure do not suggest any systematic discrimination against women in recruiting, since there are no significant differences between the percentage of women who apply and those that are hired. [16] This is consistent with another article by Bosquet, Combes, and Garcia-Peñalosa (2018) that also uses French data, and concludes that gender differences in promotion rates are mostly driven by gender differences in applications. This can be more readily seen from two graphs in Figure 1 that show the mean rates of hiring by field against the pre-reform share of women in committees and the percentage of female applicants per field. There is a strong positive correlation between both these variables and the share of women hired, but the relationship between female applicants and the gender of winners is much stronger.[17]

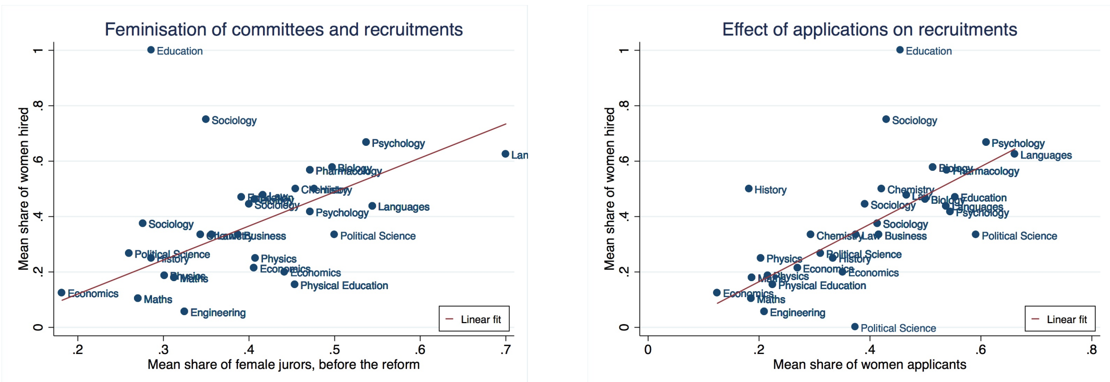


Figure 1: Determinants of hiring

Next, we can look at the effect of the reform on the share of women in committees. This reform affected fields in different ways. I present the average female/male juror ratio in Table 3 for each discipline before and after the reform. We can see a clear increase in the proportion of women sitting on committees due to the reform, except for the disciplines where the quotas were already respected, such as Biology or Languages.

[16] I cannot rule out that the coefficient from regressing whether a woman was ranked first on the share of women applicants or ranked is 1.

[17] In Table 5 which directly tests for the effect of these two variables, the positive effect of the feminisation of committee becomes insignificant once we control for percentage of female applicants.

Table 2: Descriptive statistics on ranked candidates, hires and applicants

	% of female applicants	% of women hired	% of women ranked
Biology	0.54	0.55	0.57
Business	0.42	0.27	0.55
Chemistry	0.39	0.46	0.47
Economics	0.26	0.16	0.25
Education	0.63	0.54	0.65
Engineering	0.23	0.071	0.20
History	0.28	0.36	0.31
Languages	0.56	0.47	0.63
Law	0.46	0.42	0.44
Maths	0.19	0.15	0.21
Pharmacology	0.53	0.54	0.64
Physical Education	0.15	0.11	0.15
Physics	0.20	0.23	0.18
Political Science	0.35	0.24	0.43
Psychology	0.62	0.61	0.66
Sociology	0.43	0.46	0.50
Total	0.40	0.37	0.44

This table excludes 72 committees with missing data on applications. Percentages represent averages over all committees in a field.

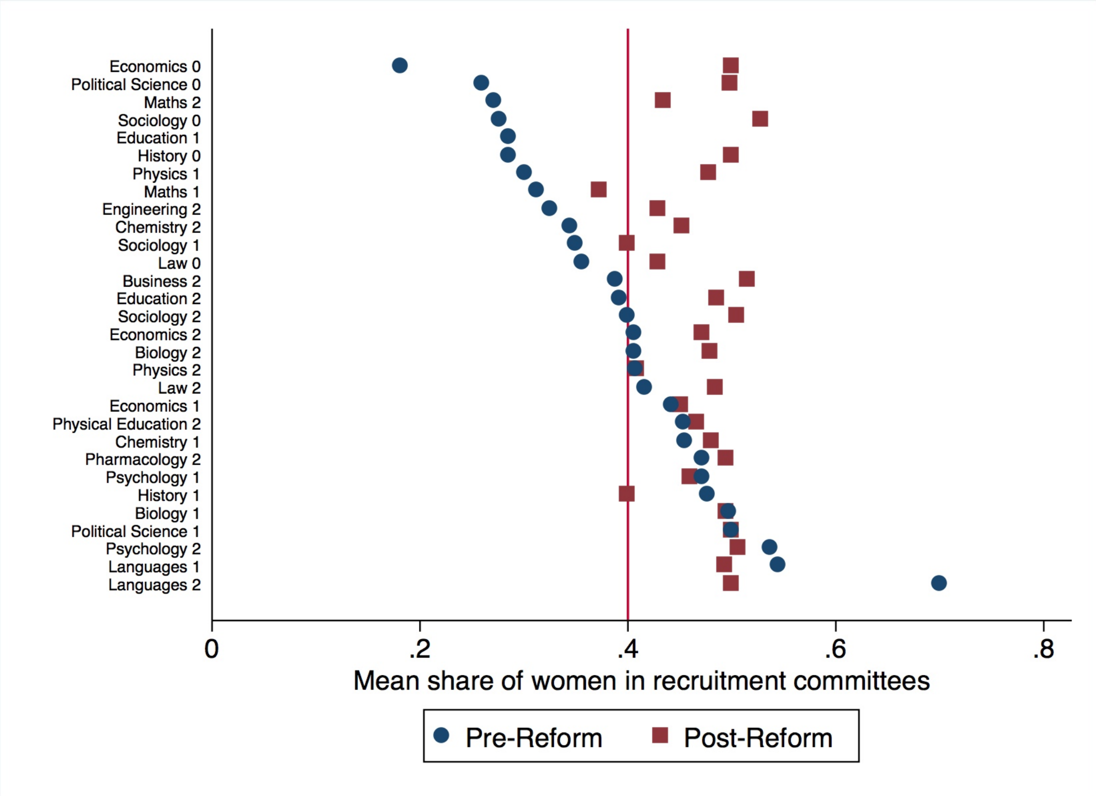
There is also some variation across universities, since some disciplines are more feminised in some universities.

When constructing the treatment and control groups, I use variation at the field-university level, assigning to the treatment group the disciplines in universities where the average proportion of women in committees pre-reform was less than 0.4. The variation at the field university level is shown graphically in Figure 2. Though the reform has a clear effect on the share of women in committees, there is no systematic effect on the gender of the jury president however. This is similar to results in other contexts. For instance, following the implementation of boardroom quotas in France, Rebérioux and Roudaut (2017) find that the newly promoted women are less likely to hold key positions in these boardrooms than newly promoted men.

2.2.1 Additional dimensions

Although I do not observe the CVs of candidates, I can control for two of the most important components of candidate quality, publications and connections. To control for publications, I collect the h-index and citations of candidates using the Scholar H-

Figure 2: Effect of the reform on the share of women in recruitment committees



index calculator (2010), which creates h-indices and citation counts from Google Scholar. [18]These variables will be used as measures, albeit imperfect,[19] of candidate quality. This matters if, on average, ranked female candidates have smaller or larger h-indices than male candidates. Table 29 in the Appendix explores this possibility.

Academic connections are another important potential confounder (see Combes, Linnemer, and Visser (2008)). In order to get more information on potential connections between jurors and candidates, I scraped the French website, theses.fr, which gathers information on which institution French doctors received their PhD from, as well as the name of their PhD supervisor. The data is harder to collect for foreign PhD students, but I gathered information from other online sources on where they received their PhD.[20]

As we can see in Table 4, the probability of having one's PhD supervisor in the committee is small (6%) but not negligible, rising to 12% in Psychology. We can also

[18]A scientist's h-index is the h number of publications he has over h citations. I also collect citation and h-indices which are discounted by the number of co-authors and how long ago the article was published.

[19]As pointed out by Bagues and Zinovyeva (2015), uninformed jurors are likely to only consider citation or h-index measures of candidates to determine whether the candidates are qualified or not, whereas more informed jurors will rely on information that can not be observed.

[20]Most juries are made up almost exclusively of French professors, so the probability of students who have received a PhD abroad of being in a committee with their PhD supervisor is very small.

Table 3: Effect of the reform on the mean share of women jurors and number of female presidents

	Mean share of women jurors		Share of female presidents	
Discipline	Pre-reform	Post-reform	Pre-reform	Post-reform
Biology	0.46	0.48	0.30	0.43
Business	0.39	0.52	0	0.29
Chemistry	0.37	0.46	0.80	0.31
Economics	0.31	0.47	0	0.12
Education	0.37	0.49	0.25	0.14
Engineering	0.33	0.43	0	0.15
History	0.33	0.45	0.45	0.50
Languages	0.55	0.50	0.56	0.42
Law	0.38	0.47	0.21	0.05
Maths	0.30	0.41	0.31	0.29
Pharmacology	0.47	0.49	0.44	0.38
Physical Education	0.45	0.47	0.14	0
Physics	0.34	0.44	0.08	0.19
Political Science	0.31	0.50	0.27	0.29
Psychology	0.51	0.49	0.23	0.62
Sociology	0.31	0.50	0.43	0.38
Total	0.39	0.47	0.30	0.30

look at endogamy, how often candidates with a PhD from the same institute apply for a position there. Once again, this is highly dependent on fields, with candidates in Biology (around 39%) very likely to have received a PhD from the same institution where they are applying for a position. I look at gender differences in citation and h-indices in Appendix F.

Table 4: Descriptive statistics on academic connections

	# Total Candidates	Candidates w/ supervisor in jury	Candidates w/ PhD from the institution	Mean h-index
Biology	204	19	79	8.68
Business	52	1	5	2.94
Chemistry	50	2	18	14.4
Economics	76	3	14	8.25
Education	53	7	36	9.96
Engineering	59	2	14	8.05
History	41	0	7	6.51
Languages	148	17	27	2.70
Law	102	8	34	2.73
Maths	267	7	39	6.66
Pharmacology	95	5	13	8.17
Physical Education	36	1	9	5.78
Physics	107	7	36	9.96
Political Science	64	3	14	9.27
Psychology	82	10	34	5.48
Sociology	112	5	11	9.03
Total	1548	90	365	6.95

Data on candidates' PhDs is scraped from online sources.

3 Model

In the administrative data used in the paper, I observe a single ranking from each committee over all candidates. I consider that jury j has a latent utility function U_{ij} when hiring candidate i which is defined as follows:

$$U_{ij} = q_{ij} + \mu g_i g_j$$

where q_{ij} is how qualified the candidate is for a particular post, as evaluated by the committee j, and μ is a parameter that evaluates how committees rank different genders, when g_i is the gender of the candidate, and g_j the gender composition of the committee.[21] In this article I consider three different methods to recover the parameter of interest, μ, from ranked data: a random utility model, a rank-ordered logit and a conditional logit. Although the rank-ordered logit directly takes ranks into account, there is no clear method for performing IV, which is why both methods are considered in this article.

[21] Since we have ranking data, I do not consider intransitive preferences, and do not study heterogeneity of utilities among committee members.

3.1 Rank-ordered Logit

One way to estimate this model could be to regress the rank of the candidate using OLS on the independent variables. However, ranking is simply a preference ordering, and the distance between the 1^{st} and 2^{nd} rank might not be the same as between 2^{nd} and 3^{rd}. The standard method to recover parameters from ranking data in economics is the rank-ordered logit, which doesn't rely on the distance between ranks.

This method can be viewed as a multiplicative form of the traditional multinomial logit, and was proposed by Beggs, Cardell, and Hausman (1981), later refined in Hausman and Ruud (1987). The intuition is that respondents rank items based on repeated multinomial logit decisions, first choosing item i from M choices, then item i' from $M-1$ choices and so forth, with the choice set decreasing in size until only one alternative is left. Combes, Linnemer, and Visser (2008), use this method to analyse whether connections matter for academic promotions, using ranking data from the French *agrégation* in economics, a context very similar to ours.

Putting more structure on U_{ij}, consider the case where q_{ij} has the following form:

$$q_{ij} = x_{ij}\beta + \eta_j + \epsilon_{ij}$$

where x_{ij} is a vector with the characteristics of the candidates, η_j is a jury fixed effect and ϵ_{ij} is an error term. Consider two variables, q^*_{ij}, which is equal to the rank k given by jury j to alternative i, and r_{ij}, which is equal to the alternative i given the rank k by jury j. If ϵ_{ij} is distributed Type 1 extreme value, so that $Pr(\epsilon_{ij} < u) = e^{-e^{-u}}$, then the likelihood of observing a particular ranking in the j^{th} concours is:

$$Pr(U_{r_{1j}j} > U_{r_{2j}j} > \ldots > U_{r_{Mj}j})$$

$$l_j(\beta) = \prod_{i:q^*_{ij}=1}^{M-1} \frac{\exp(x_{ij}\beta + \mu g_i g_j)}{\sum_{i':ri'j \geq rij}^{M} \exp(x_{i'j}\beta + \mu g_{i'} g_j)}$$

Notice here that the jury fixed effects, η_j cannot be recovered, since they do not enter the likelihood function. We can only recover parameters that vary for candidates within jury. The log-likelihood function of J independent committees is then:

$$\sum_{j=1}^{J}\sum_{i:q^*_{ij}=1}^{M-1} x_{ij}\beta + \mu g_i g_j - \sum_{j=1}^{J}\sum_{i:q^*_{ij}=1}^{M-1}(\log \sum_{i':ri'j\geq rij}^{M} \exp(x_{i'j}\beta + \mu g_{i'} g_j))$$

The model can be estimated through maximum likelihood. Since the log likelihood is globally concave, β and μ will be the unique maximisers. We can recover the parameters μ by considering (for instance) the effect on the rank of women (relative to men) of having a female president (or a committee made up of many female members) or the effect for both genders of having a president of the same gender. In the estimation, I also use a conditional logit to estimate the probability of being first ranked. The reasoning is very much the same as the one above, except that the conditional logit can answer the following question: What is the probability of ranking candidate i first from a pool of M candidates.

3.2 Random utility model

The first models using a latent utility framework date back to Thurstone (1927). In those early models, ranked data was analysed by pairwise difference. A part of the estimation will be devoted to this approach, since it allows for a more flexible approach to instrumental variable regressions.

We have a latent utility function where the candidate i has a better ranking than candidate i' if

$$U_{ij} \geq U_{i'j}$$

This implies

$$q_i \geq q_{i'} - \mu g_i g_j + \mu g_{i'} g_j$$

Assume that q_i, the candidate's intrisic quality as evaluated by jury members depends on two factors: an observed term, x_i, and an unobserved term, $\epsilon_{ij} \sim N(0, \sigma)$ which represents other variables that the jury may take into account. If we represent the probability of candidate i receiving a better ranking than candidate i' by a binary variable Y, then:

$$Pr(Y = 1|x) = Pr(U(i) \geq U(i'))$$

$$Pr(x_i\beta + \mu g_i g_j + \epsilon_{ij} \geq x_{i'}\beta + \mu g_{i'} g_j + \epsilon_{i'j})$$

$$Pr(\beta(x_i - x_{i'}) + \mu(g_i - g_{i'})g_j \geq \epsilon_{i'j} - \epsilon_{ij})$$

$$= \Phi(X\beta + \mu G g_j)$$

where $X = x_i - x_{i'}$ and $G = g_i - g_{i'}$. We can then use a probit regression to recover the parameters. Of course, when $g_i = g_{i'}$, $G = 0$, so we can only estimate μ on pairs of candidates of different gender. Only dyads with candidates of different gender can give us information on own-gender preferences.[22]

4 Estimation

In this section I estimate the models from Section 3 on my dataset. I look at results without taking into account possible endogeneity of assignment of candidates to committees, before analysing the effect of the reform in Section 5.

In the first table I look at the effect of two variables, the gender of the committee president and the proportion of females in the committee on the probability that the first-ranked candidate is a woman, using a probit regression. Since in our dataset, most candidates who are ranked first accept the offer, this tests directly for whether more women are hired when there are more women in the committee. The results are in Table 5. The first column has no controls, and there is a significant, positive effect of having more women in the committee in this case. However, once we control for the percentage of female candidates, this effect vanishes, and the coefficient becomes negative and insignificant. I drop committees where only men or women applied in column 3, and add discipline fixed effects and a dummy for whether the post is for an assistant or associate professorship.[23] The gender of the committee president is insignificant in all columns.

This methodology however might not be the most appropriate. One problem is that

[22]Simulations, presented in Section B show that these estimates are unbiased for the data generating process presented above, and that the standard errors from OLS (if the parameter values are not too large) or probit regressions have the right power once we cluster at the committee level, to take correlation of errors (since we may use the same individual from the same committee multiple times in our sample) into account.

[23]Maitre de Conférences or Professeur des Universités in French.

Table 5: Correlation between the gender of jurors and the probability of women being first-ranked

Dependent variable:	Gender(W) of the first ranked candidate			
Estimator:	Probit	Probit	Probit	Probit
Marginal effects	(1)	(2)	(3)	(4)
Female president	0.010	-0.015	-0.039	-0.053
	(0.05)	(0.04)	(0.05)	(0.05)
Share of women in the committee	0.733***	0.020	0.120	0.118
	(0.19)	(0.17)	(0.24)	(0.25)
Share of female candidates		0.909***	0.780***	0.858***
		(0.05)	(0.16)	(0.17)
Professorship				0.094*
				(0.06)
Discipline fixed effects	No	No	Yes	Yes
University FE	No	No	Yes	Yes
# observations	455	455	359	359

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Contests with only members of a specific gender are dropped in columns three and four.

we cannot control for variables at the candidate level. Even if gender had no effect on the committee's decision, we could still find negative effects if better male candidates apply to committees which are mostly composed of women. A solution to this problem is to estimate the probability of being hired through a conditional logit. A conditional logit estimates the probability that candidate i is picked out of a pool of applicants. The results in the first three columns of Table 6 can be interpreted as the effect of the feminisation of committees on the likelihood that a woman is first ranked. Please note that for all variables estimated at the committee level, the results presented are the differential effects for women with respect to men. The correlations here are negative and significant, (i.e. more feminised committees are less likely to hire women) but controlling for field fixed effects and connections make the coefficient statistically insignificant, though the point estimate remains high.

Although we can likely rule out a large positive correlation of the feminisation of committees on the probability of women being hired, this doesn't preclude finding an effect on the *ranking* of women. I directly analyse ranks using the rank-ordered logit, in columns 4-6 of Table 6. With this method I can only control for variables that affect ranking within committees. For all non-individual variables, I consider their effect on the rank of women relative to men. Similar to the results above, we cannot estimate these parameters in committees where only men or only women applied. Therefore, we have

Table 6: Correlation between the gender of jurors and the gender of first ranked candidates

Dependent variable:	Ranked first			Rank		
Estimator:	C. Logit	C. Logit	C. Logit	RO. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)	(5)	(6)
Share of women in committee		-0.710**	-0.874		-0.422**	-0.126
		(0.35)	(1.30)		(0.20)	(0.78)
Female president		-0.357	-0.526		-0.040	-0.081
		(0.29)	(0.32)		(0.16)	(0.18)
Standardised age-adjusted H-index	0.326***	0.320***	0.354***	0.275***	0.275***	0.286***
	(0.07)	(0.07)	(0.08)	(0.04)	(0.04)	(0.05)
Candidate is a woman	-0.417***		-0.569	-0.212***		-0.286
	(0.14)		(0.80)	(0.08)		(0.42)
Share of female candidates			0.360			-0.290
			(0.84)			(0.44)
PhD supervisor in the committee			0.474*			0.350**
			(0.26)			(0.15)
PhD from the same institute			0.580***			0.296***
			(0.17)			(0.10)
Discipline Fixed Effects	No	No	Yes	No	No	Yes
University Fixed Effects	No	No	Yes	No	No	Yes
# observations	1357	1357	1357	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Committees where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

less observations than in the descriptive statistics presented in section 2.

A first interesting result is to look at the effect of being a woman on the rank given by the committee for both specifications, in columns 1 and 4. This effect is negative and significant even when controlling for the h-index.[24] Though this article does not focus on discrimination *per se*, this column is the closest to a direct test since it asks whether women are as likely to be hired as men with the same publication record within a given committee. Next, consider the interaction between how women are ranked, and the gender composition of the committee. There is a significant and large negative correlation between the share of women in the committees on the ranking of women in columns 2 and 3, where the only control is the h-index of candidates. However, once more control variables are added, this effect becomes insignificant, although the point estimate remains high,especially in the conditional logit specification. One way to interpret the estimate is to compare it to the coefficient that controls for publications one row below. In column 3 for instance, the effect of having a jury made up of 10% more women in this case is comparable to a 0.25 standard deviation drop in h-index.

[24] I use the age-adjusted h-index of candidates since this is the publication measure that is the most significant when including all other measures, as seen in Table 32 in the Appendix. It is standardised by field and seniority since the average h-index varies greatly along these dimensions

Table 7: Regression of the gender composition of applicant pools on the gender composition of the recruitment committee

Dependent variable:	Gender composition of applicant pool			
Marginal effects	(1)	(2)	(3)	(4)
Share of women in the committee	0.691***	0.548***	0.212**	0.093
	(0.11)	(0.09)	(0.10)	(0.07)
Female president	0.031	0.007	0.005	-0.014
	(0.03)	(0.02)	(0.02)	(0.02)
Discipline fixed effects	No	No	Yes	Yes
# observations	455	359	455	359

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Standard errors are clustered at the jury level. Columns 2 and 4 drop contests with only male or female applicants.

The control variables here are significant and have the expected sign: having a supervisor in the committee, or having a PhD from the institute where you are applying increases your ranking by the committee, and the effect size here is larger than a 1 standard deviation increase in the h-index. This result confirms previous research on the importance of networks and connections in academia.[25]. Results using the random utility model are very similar and are presented in the Appendix.

4.1 Investigating endogeneity concerns

One important issue is that the selection of candidates into the candidate pool is possibly endogenous. For instance, candidates may put less effort in their application depending on the composition of the jury, or even decide not to apply at all. I investigate this concern by considering whether the composition of the jury affects variables other than the ranking within field. In the following tables, I present results both with contested and uncontested (only candidates of a single gender applied) committees. In Table 7, I regress the gender ratio of candidates on the gender ratio of jury members, while controlling for discipline fixed-effects.

The applicant pool depends significantly on the gender composition of the committee in three of the estimations, either through sub-discipline effects (i.e. sub-disciplines within each field have different gender compositions) or through endogenous selection of candidates (candidates observe the gender composition of the committee and then decide whether to apply or not), though the effect becomes insignificant once I control for

[25]See Combes, Linnemer, and Visser (2008) or Colussi (2018) for examples of connections influencing promotions and publications

discipline fixed effects and drop uncontested committees. We may therefore be worried that the results presented above could be endogenous, since candidate pools may have selection bias on unobservable candidate quality. In the following section, I try to solve this endogeneity concern by using variation from a policy reform.

5 Effect of the quota

In this section, I look at the effect of the gender quota reform of 2015, which is used as an exogenous shock to analyze the effect of the share of women in the committee on the hiring and ranking of women. I use a diff-in-diff for the conditional and rank-ordered logit estimations, and an instrumental variable regression using a linear probability model[26] for the random utility model.

I use the reform to construct a control and a treatment group: I assign fields in universities where the gender composition of the jury is below 40% before the reform in 2015 to the treatment group. For instance, if on average, committees in Law in University A are made up of 35% of women on average before 2015, and of 45% on average in University B, then committees in Law in University A will be assigned to the treatment group, and Law committees in University B will be assigned to the control group. In practice, most fields are consistently assigned to treatment or control groups across universities.

We can first consider whether the reform had any effect on hiring. Are women more or less likely to be first-ranked in the treatment group after the reform? I instrument for the percentage of women in the committee using the effect of being in the treatment group post-reform. The results are presented in Table 8. There is a large negative point estimate of having more women in recruitment committees on the probability that a woman is hired, but this effect is insignificant. This effect is comparable in size to the one of having more women candidates. Note that I cannot control for individual specific characteristics in this specification, which may explain why the standard errors on this coefficient is so large.

Using the conditional logit specification however paints a different picture. The spec-

[26]The coefficients from probit regressions are biased whenever the errors are heteroscedastic, which could be the case here since the instrument is binary. I present results using probit regressions in Appendix F.

Table 8: IV on gender of first-ranked candidate

Dependent variable:	Gender of first-ranked candidate		
Estimator:	IV	IV	IV
	(1)	(2)	(3)
Share of women in the committee	-0.616	-0.803	-1.080
	(0.68)	(0.59)	(0.80)
Treatment Group	-0.260***	-0.089	-0.107
	(0.08)	(0.06)	(0.08)
Reform	0.104	0.081	0.092
	(0.07)	(0.06)	(0.07)
Female president		-0.003	-0.001
		(0.05)	(0.06)
Share of female candidates		1.001***	1.015***
		(0.07)	(0.17)
University fixed effects	No	No	Yes
# observations	455	455	359

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. This regression considers first-ranked candidates only. Contests with applicants from a specific gender only are dropped in column three. Instrumental variable here is being in the treatment group post-reform.

ification estimated is the following:

$$\mathrm{F}_{ij} = \alpha R_j T_j g_i + \gamma T_j g_i + \delta R_j g_i + \beta_1 X_i + \beta_2 g_i Y_j + \epsilon_{ij}$$

where F_{ij} is a variable equal to 1 if the candidate is ranked first, and 0 if not, γ represents the effect of being in the treatment group before the reform, δ the effect of the reform on the control group, and α is the parameter of interest, the effect of being in the treatment group after the reform relative to the control group. g_i is a binary variable equal to 1 if the candidate is a woman, X_i is a vector of individual characteristics of the candidate (such as the candidate's h-index or whether the candidate received his PhD from the university he is applying to), and Y_j a vector of committee characteristics (such as the share of female applicants) that are interacted with g_i. The model is then estimated using a conditional logit.

The results are below in Table 9. We can see that the reform has a large, significant and negative effect on the hiring of women. This effect is robust to the inclusion of other covariates and field fixed effects. Connexions and publications have a positive effect on hiring, as expected. The results can be compared to the size of the coefficient on the H-index below. The negative effect of the reform is comparable to a 4 sd deviation drop in the h-index of candidates. However, this seems to be partially driven by the large

Table 9: Effect of the quota on the first ranked candidates

Dependent variable:	Ranked first		
Estimator:	C. Logit	C. Logit	C. Logit
	(1)	(2)	(3)
Post Reform · Treatment Group	-1.525***	-1.521***	-1.671***
	(0.56)	(0.56)	(0.64)
Post-reform	0.753**	0.738*	0.858**
	(0.38)	(0.38)	(0.40)
Treatment Group	0.850**	0.806*	0.045
	(0.41)	(0.43)	(0.65)
Female president	-0.556*	-0.475	-0.678**
	(0.32)	(0.32)	(0.34)
Candidate is a woman	-0.688**	-0.488	-0.766
	(0.29)	(0.51)	(0.85)
Standardised age-adjusted H-index		0.323***	0.359***
		(0.07)	(0.08)
Share of female candidates		-0.327	0.535
		(0.76)	(0.84)
PhD supervisor in the committee			0.477*
			(0.27)
PhD from the same institute			0.587***
			(0.17)
Professorship			0.930***
			(0.32)
Discipline fixed effects	No	No	Yes
# observations	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Treatment variable here is based on mean pre-reform gender parity at the field*university level. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Committees where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

positive effect of the reform on the control group. [27]

Next, we move to the effect of the reform on ranks. To test for the effect of the reform, I estimate the equation below:

$$\text{Rank*}_{ij} = \alpha R_j T_j g_i + \gamma T_j g_i + \delta R_j g_i + \beta_1 X_i + \beta_2 g_i Y_j + \epsilon_{ij}$$

where Rank* is the latent variable that the members of the jury use to rank candidates. This specification is otherwise very similar to the conditional logit. The model is estimated through rank-ordered logit.

The results using the rank-ordered logit are presented in Table 10, which is the main

[27]This is slightly worrying, but also consistent with the IV estimates below that show the control group actually has a reduction in the share of women candidates in the first stage. The rank-ordered logit estimates, which use the data more efficiently, show no significant effect in the control group.

table of this paper. As we can see, there is a strong negative effect on the rank of women of being in a field that had a large increase in the share of women jurors due to the reform.[28] There is no significant effect on the control group in this specification. This effect is robust to the inclusion of other covariates, such as the standardised h-index of candidates or the effect of having a PhD from the same institute. To interpret the size of the effect, one can compare the coefficient in the first row with the one for the standardised h-index. In column 2 for instance, the size of the effect is equivalent to a 4 standard deviation drop in the h-index of candidates. [29]

Another way to interpret this result is to think about the women that weren't hired as a result of the reform, which we can do using some back-of-the-envelope computations. I can simulate the predicted probabilities from the rank-ordered logit to recover a predicted rank, and in particular a predicted first-ranked candidate. We can then analyse the effect of the quota under two counterfactual scenarios: One in which there is no effect of the reform in the treatment group (i.e. $\alpha = 0$ and δ=0 if the contest is in the treatment group), and one in which there is no reform at all (i.e. α and δ=0). These scenarios are presented in Table 11. Relative to the true number of women recruited, the number of women recruited if there had been no effect of the reform on the treated would have increased by 38%. If there had been no effect of the reform on the treated and control groups, we would still have expected an increase of 15% in the number of women recruited.

What do these results imply for the effect size of increasing women in recruitment committees? We can answer this question by using the dyads, which allow us to estimate the model through IV, instrumenting for the potentially endogenous variable with the effect of being in our group post reform. I estimate the following equations:

$$W_j = \alpha^1 R_j T_j + \gamma^1 T_j + \delta^1 R_j + \beta^1 (X_i - X_{i'}) + \beta_2^1 Y_j + \epsilon_{ij}^1$$

$$\mathrm{D}_{ii'j} = \mu \hat{W}_j + \gamma^2 T_j + \delta^2 R_j + \beta_1^2 (X_i - X_{i'}) + \beta_2^2 Y_j + \epsilon_{ij}^2$$

where $D_{ii'j}$ is a binary variable equal to 1 if the female candidate i has a better rank

[28] The standard errors presented are not clustered at field level. In fact clustering at discipline level or the discipline-university level actually dramatically lowers the standard errors for almost all the estimations, and the results in Table 14 for instance become significant at the 5% level when clustering the standard errors.

[29] I find no evidence of pre-trends, as can be seen in Table 36 in the Appendix

Table 10: Effect of the quota on the rank of female candidates

Dependent variable:	Rank		
Estimator:	RO. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)
Post Reform · Treatment Group	-0.808***	-0.793***	-1.082***
	(0.31)	(0.31)	(0.41)
Post-reform	0.304	0.283	0.226
	(0.20)	(0.20)	(0.23)
Treatment Group	0.401*	0.354	-0.166
	(0.22)	(0.22)	(0.30)
Female president	-0.082	-0.063	-0.145
	(0.16)	(0.17)	(0.17)
Candidate is a woman	-0.351**	-0.110	-0.044
	(0.14)	(0.25)	(0.41)
Standardised age-adjusted H-index		0.274***	0.286***
		(0.04)	(0.05)
Share of female candidates		-0.392	-0.219
		(0.40)	(0.43)
PhD supervisor in the committee			0.368**
			(0.16)
PhD from the same institute			0.316***
			(0.10)
Professorship			0.390**
			(0.18)
Discipline fixed effects	No	No	Yes
# observations	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Treatment variable here is based on mean pre-reform gender parity at the field*university level. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

Table 11: Counterfactual Recruitments

	True number	Simulated number	No reform for treated	No reform for treated and control
Man hired	118	128	88	106
Woman hired	78	68	108	90
Total	196	196	196	196

These results show the gender of the true and simulated first-ranked candidates in the post-reform period under different scenarios. Column 1 represents the true numbers of men and women recruited, column 2 the gender of the simulated first-ranked candidates from the estimates in column 3 of Table 10. Column 3 represents the gender of the simulated first ranked candidates but imposes no effect of the reform for the treated group. Column 4 imposes no effect for both treatment and control group.

than the male candidate i' within a dyad, $(X_i - X_{i'})$ is a vector of differences between candidate level characteristics, for instance the difference in h-index between candidates i and i', $\hat{W}$ is the instrumented share of women in committees, and μ is our coefficient of interest, estimating the effect of having more women in recruitment committees on the probability of being hired. The model is estimated through 2SLS.

The first-stage estimates are presented in Table 12.[30] Being in the group affected by the reform has a highly significant and positive (around 15 percentage points) effect on the share of women in committees, as expected. There is also a negative effect of being in the control group on this share, which is significant at the 10 percent level. The instrumental variable results are presented in Table 13.[31] The share of women in the committee has a significant negative effect. The effect size can now be interpreted in terms of changes in the share of women in committees. The point estimates for column 2 imply that a 10% increase in the percentage of women in committees has the same effect on the probability that a woman is more highly ranked in a dyad than a 2.5 standard deviation decrease in the h-index.

The results can be considered as a Local Average Treatment Effect (LATE) i.e. the effect of increasing the share of women jurors in disciplines that have few women. These effects may be stronger or weaker than the ones we would find when analysing all fields. However, the measure that we are considering is the relevant one, since the policy proposals that are usually debated aim at increasing the proportion of women evaluators in fields where they are not represented, not in fields where parity is already achieved. Whether increasing the proportion of women evaluators also increases the rank of women in fields where parity is already respected is a moot point. In fact, the global effect of the reform, presented in Appendix E, is close to 0.

[30] The F-test statistic for the first-stage is above 10 (and in fact close to 30) in all columns, as advised by Staiger and Stock (1997), and the instrument is significantly correlated with the endogenous variable.

[31] Of course, one could simply look at the probability of a woman being better ranked than a man after the reform in the disciplines affected, using a difference-in-difference approach rather than instrumental variables. Table 26 in the Appendix provides results using this methodology.

Table 12: First stage: IV

Dependent variable:	Rank		
Estimator:	IV	IV	IV
	(1)	(2)	(3)
Post Reform · Treatment Group	0.155***	0.155***	0.153***
	(0.03)	(0.03)	(0.03)
Treatment Group	-0.179***	-0.178***	-0.156***
	(0.03)	(0.03)	(0.02)
Reform	-0.032*	-0.031*	-0.033*
	(0.02)	(0.02)	(0.02)
Female president	0.030**	0.030**	0.031**
	(0.01)	(0.01)	(0.01)
Standardised age-adjusted H-index difference		-0.004	-0.004
		(0.00)	(0.00)
PhD supervisor in the committee			-0.003
			(0.01)
PhD from the same institute			-0.002
			(0.01)
Share of female candidates			0.109***
			(0.03)
Professorship			-0.024
			(0.01)
F-Statistic	32.00	32.27	32.91
# observations	1086	1086	1086

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man.

Table 13: IV estimate of the effect of the increase in women jurors on the probability that a woman is better ranked within a dyad

Dependent variable:	Rank		
Estimator:	IV	IV	IV
	(1)	(2)	(3)
Share of women in the committee	-1.858***	-1.856***	-2.370***
	(0.65)	(0.64)	(0.81)
Female president	0.025	0.023	0.033
	(0.05)	(0.05)	(0.06)
Treatment Group	-0.174**	-0.173***	-0.431***
	(0.07)	(0.06)	(0.15)
Reform	0.078	0.075	0.039
	(0.05)	(0.05)	(0.06)
Standardised H-index difference		0.072***	0.074***
		(0.01)	(0.01)
PhD supervisor in the committee		0.059	0.069
		(0.05)	(0.06)
PhD from the same institute		0.045	0.053
		(0.03)	(0.04)
Share of female candidates			0.095
			(0.23)
Professorship			0.028
			(0.06)
Discipline fixed effects	No	No	Yes
# observations	1086	1086	1086

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The gender variable is coded as 1 for a woman and 0 for a man. Errors are clustered at the contest level.

5.1 Continuous specification

Another way of approaching the problem is to consider a continuous specification of the treatment variable. Using a binary variable to determine assignment to treatment, though it enables an easy interpretation of the treatment variable, loses some of the information since fields that were far away from the cut-off before the reform are weighted in the same away as those that were close. A standard way to estimate the effect of treatment intensity would be to take the difference in the outcome variable post- and pre-reform, and then regress this new variable on the distance to the threshold pre-reform.[32] It is impossible to do this using my data however since my observations are at the field level, and I therefore cannot difference my outcome variable unless I collapse the data at field or field university level. I use the following specification instead to estimate the effect of the reform on ranks of women:

$$\mathrm{DV}_{ij} = \xi g_i \max\{0, 0.4 - \mathrm{jury}_x^{pre-2015}\} + \mu_{ij} g_i^{post-2015} \max\{0, 0.4 - \mathrm{jury}_x^{pre-2015}\}$$

$$+\beta_1 X_i + \beta_2 g_i Y_j + \epsilon_{ij}$$

where ξ captures the effect of the treatment intensity before the reform and jury is the mean share of female jurors in a field before 2015, and DV is the dependent variable of interest, the rank or whether the candidate finished first depending on the specification. Our parameter of interest here is μ_{ij}, which should capture the post-reform effect of the treatment. The results for the conditional logit and the rank-ordered logit can be seen in Table 14, and are similar to those presented in the previous tables, though the results for the ranks here are fragile and significant at the 10 percent level only. An interpretation of the results presented here is that being in a field that was 10 percentage points away from the cut-off before the reform has an effect similar to a 3 unit decrease in age-adjusted h-index.

[32] See Casas-Arce and Saiz (2015) for an example

Table 14: Effect of the quota on the rank of female candidates-Continuous specification

Dependent variable:	Ranked first		Rank	
Estimator:	C. Logit	C. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)
Distance to threshold-post reform	-13.022**	-14.672**	-6.320*	-6.770*
	(5.94)	(6.93)	(3.55)	(4.05)
Distance to threshold-pre reform	6.556	9.426	4.746	2.651
	(5.23)	(10.38)	(3.07)	(4.90)
Post-reform	0.499	0.750*	0.096	0.106
	(0.35)	(0.40)	(0.20)	(0.24)
Female president	-0.426	-0.587	-0.040	-0.092
	(0.31)	(0.36)	(0.16)	(0.19)
Standardised age-adjusted H-index	0.323***	0.357***	0.278***	0.290***
	(0.07)	(0.08)	(0.05)	(0.05)
Candidate is a woman	-0.501	-1.691	-0.748	-0.732
	(0.85)	(1.53)	(0.48)	(0.72)
Share of female candidates	-0.449	0.401	-0.216	-0.220
	(0.78)	(0.86)	(0.40)	(0.45)
PhD supervisor in the committee		0.422		0.375**
		(0.26)		(0.16)
PhD from the same institute		0.581***		0.309***
		(0.18)		(0.10)
Professorship		0.896***		0.351*
		(0.34)		(0.19)
Discipline fixed effects	No	Yes	No	Yes
University fixed effects	Yes	Yes	Yes	Yes
# observations	1357	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables such as discipline fixed effects must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

Table 15: Effect of the reform on the candidate pool

Dependent variable:	Gender composition of candidates					
Marginal effects	(1)	(2)	(3)	(4)	(5)	(6)
Reform	0.013	0.067**	0.019	0.036	0.009	0.033
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Female president	0.049	-0.018	0.016	0.065	-0.044	0.021
	(0.04)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)
Professorship	-0.166***	-0.117***	-0.146***	-0.188***	-0.090***	-0.147***
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.03)
Treatment Group			-0.023			0.058
			(0.06)			(0.08)
Treatment Group*Reform			0.045			-0.026
			(0.05)			(0.05)
Group	Control Group	Treatment Group	Both	Control Group	Treatment Group	Both
Discipline Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# observations	235	220	455	166	130	296

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The last 3 columns are restrained to 2013-2016.

5.2 Effect of the reform on the candidate pool

A potential effect of the reform is to encourage more women to apply for professorial positions, if they believe that more feminised committees will be better inclined towards them. This could bias our results if the composition of the applicant pool changes drastically. To check whether the approach used to estimate the effect of the reform is valid, I test whether the gender composition of the applicants changes significantly before and after the reform in both groups. There are in fact changes in the gender composition of the applicant pool after the reform, though as shown in column 3, there are no significant differences between treated and control group in this respect. This could also be due to the long time span in my data, with committees from 2009 to 2018.

One way to investigate this problem is to restrict the time span to the period around the reform, from 2013 to 2017. In this time frame, there are no significant changes in the gender composition of the applicant pool after the reform in both treated and control groups, and as seen in column 6 the point estimate for the difference between treated and control group is in fact negative. This indicates that the changes in the applicant pool seen in columns 1 and 2 could be due to longer-run trends in the propensity of women to apply for these positions rather than the reaction of candidates to the reform. We can conclude that the reform doesn't seem to encourage more women to apply to professorial positions, at least in the very short-term.

Nevertheless, Table 16 presents results using the years immediately leading up to and

Table 16: Effect of the quota on the rank of female candidates: Years 2013-2016

Dependent variable:	Ranked first		Rank	
Estimator:	C. Logit	C. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)
Post Reform · Treatment Group	-1.931***	-2.366***	-0.882**	-1.279***
	(0.70)	(0.79)	(0.41)	(0.47)
Post-reform	0.660	0.921*	0.200	0.291
	(0.46)	(0.52)	(0.24)	(0.26)
Treatment Group	1.323**	0.905	0.824***	0.320
	(0.53)	(1.12)	(0.32)	(0.55)
Female president	-0.220	-0.464	0.166	-0.073
	(0.41)	(0.53)	(0.23)	(0.26)
Candidate is a woman	-0.699	-1.016	-0.404	-0.254
	(0.68)	(1.30)	(0.32)	(0.57)
Standardised age-adjusted H-index	0.406***	0.453***	0.301***	0.320***
	(0.08)	(0.09)	(0.05)	(0.06)
Share of female candidates	-0.246	0.710	-0.138	0.154
	(1.04)	(1.23)	(0.52)	(0.54)
PhD supervisor in the committee		0.534*		0.281
		(0.32)		(0.19)
PhD from the same institute		0.753***		0.333***
		(0.22)		(0.12)
Professorship		0.833*		0.514**
		(0.43)		(0.25)
Discipline fixed effects	No	Yes	No	Yes
# observations	926	926	926	926

$*p < 0.10, **p < 0.05, ***p < 0.01$. Only the four years each side of the reform are included in this regression. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

following the reform to alleviate this concern. I present the results with this limited time frame, only including the years from 2013 to 2016. The point estimates are once again very close to those presented in the other tables, so we can probably rule out the idea that the estimated effect of the reform is due either to longer-run trends, or due to the changing nature of the application pool as a reaction to the reform.

5.3 Top ranks only

One concern in using ranked data is that some committees may not care about how they rank lower quality candidates, since they are very unlikely to be hired. To investigate this concern, I present results in Table 17 using again the conditional and rank-ordered logit specifications, but keeping only the top 3 candidates from each committee, and dropping

Table 17: Rank-Ordered Logit using the quota: top 3 ranks only. 2013-2016

Dependent variable:	Ranked first		Rank	
Estimator:	C. Logit	C. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)
Post Reform · Treatment Group	-1.841**	-2.201***	-1.528***	-1.931***
	(0.73)	(0.84)	(0.53)	(0.62)
Post-reform	0.624	0.965*	0.103	0.402
	(0.51)	(0.58)	(0.39)	(0.45)
Treatment Group	1.039*	0.901	0.937**	0.957
	(0.57)	(1.16)	(0.41)	(0.97)
Female president	-0.526	-0.658	-0.355	-0.443
	(0.44)	(0.58)	(0.32)	(0.36)
Candidate is a woman	-0.526	-1.010	-0.161	-0.812
	(0.75)	(1.38)	(0.51)	(1.01)
Standardised age-adjusted H-index	0.337***	0.380***	0.267***	0.275***
	(0.09)	(0.11)	(0.09)	(0.09)
Share of female candidates	-0.097	0.951	-0.273	0.360
	(1.13)	(1.43)	(0.81)	(0.96)
PhD supervisor in the committee		0.846**		0.505
		(0.42)		(0.36)
PhD from the same institute		0.846***		0.629***
		(0.25)		(0.19)
Professorship		0.673		0.344
		(0.44)		(0.33)
Discipline fixed effects	No	Yes	No	Yes
# observations	597	597	597	597

$*p < 0.10, **p < 0.05, ***p < 0.01$. Only the first three ranked candidates are included. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

candidates that were auditioned and not ranked. I keep the two years either side of the reform only as in Table 16. In this specification, the reform still has a significant effect and the point estimates are even larger than those that consider the full ranking. We can therefore rule out that the effect of the reform is driven by inattention or lack of effort from juries when ranking candidates that are very unlikely to be hired.

5.4 Other dimensions of the reform

In this section, I analyse the effect of the reform on other dimensions than the recruitment of women. A first question is related to how committees are formed as a reaction to the quota. Are the new jury members coming from inside or outside the department that is recruiting, taking into account the fact that at least half the members of the

Table 18: Increase in the share of internal and external members

Dependent variable:	Share of women among internal members	Share of women among external members
Estimator:	OLS	OLS
	(1)	(2)
Post Reform·Treatment Group	0.174***	0.090**
	(0.03)	(0.04)
Treatment Group	-0.167***	-0.177***
	(0.03)	(0.03)
Reform	-0.012	0.017
	(0.02)	(0.03)
# observations	455	455

$*p < 0.10, **p < 0.05, ***p < 0.01.$

committees have to be from outside the department? In Table 18, I regress the share of women among internal members and among internal members on the treatment group and reform variables, in columns 1 and 2 respectively. In both cases, the effect of being affected by reform is positive, though the point estimate for internal members is larger.[33] We can conclude that as a reaction to the reform, committees increase both the number of external and internal female committee members, though they seem to draw more from their colleagues inside the department. This result implies that women inside the departments most affected by the reform have more administrative work post-2015 relative to their male colleagues, which could harm their ability to publish.

Does the reform affect the h-index of committee members? We could think that the new female jurors have less publications than the male professors that they replace. In Table 19, I regress the mean h-index of jurors by committees on the treatment and reform dummies. There doesn't appear to be much of an effect of the reform, with the negative treatment effect on the mean h-index of jurors both very small (less than 0.05 of a standard deviation in the h-index) and insignificant. Combined with the results on internal versus external members, this shows that two concerns we could have about the reform; that jury president choose junior women on purpose to negate the effects of the reform, and that the new female jurors are less qualified than the members they replace, degrading the overall "quality" of the committee, do not seem to hold in the data.

We can also look at other jury characteristics, such as the average h-index of jury members and the age and h-indices of the jury presidents [34], and see how these variables

[33]Note that since the IV estimate is essentially a Wald estimator, using the IV strategy from above to instrument for the internal and external share would necessarily give a larger effect size for the group that was least affected by the reform.

[34]For a few presidents I was not able to recover the age of the presidents. This leads to the change in the number of observations

Table 19: Effect of the reform on average h-index of jury members

Dependent variable:	Mean standardised h-index of jurors	Mean standardised h-index of female jurors	Mean standardised h-index of male jurors
Estimator:	OLS	OLS	OLS
	(1)	(2)	(3)
Treatment Group	0.055	0.034	0.066
	(0.06)	(0.08)	(0.10)
Reform	-0.002	-0.005	0.013
	(0.05)	(0.07)	(0.08)
Post Reform*Treatment Group	-0.050	-0.033	-0.055
	(0.07)	(0.10)	(0.12)
Professorship	0.227***	0.457***	0.352***
	(0.04)	(0.05)	(0.06)
University Fixed Effects	Yes	Yes	Yes
# observations	455	455	455

$*p < 0.10, **p < 0.05, ***p < 0.01$.

affect the probability of a woman being hired. I estimate 4 models in Table 18 that include these variables into the rank-ordered and conditional logit. In the standard model that does not look at the effect of the reform, none of these variables have a significant effect. Two variables are significant at the 10% level in the rank-ordered logit however once I include these variables in the model that looks at the effect of the quota. Older presidents, and committees where the male committee members have high h-indices seem to favour women. [35]

6 Discussion - How should we interpret this result?

I provide a graphical interpretation of the results with the instrument in Figure 3. In this figure, I plot the coefficients from a rank-ordered logit of being a woman, in the two years before and after the reform, controlling for the h-index of candidates. The treatment group consists of recruitments in disciplines that are affected by the reform, while the control group is made up of disciplines that were already respecting gender parity in committees, and should therefore not be affected by the reform, as explained above. Slightly surprising is that pre-reform, fields that were not feminised actually tended to favour women. In disciplines affected by the reform, the rank of women decreases, whereas women are ranked slightly higher in the control group. Figure 4 plots the difference between control and treatment group, with the standard errors on the difference.

This effect does not seem to be driven by outliers in fields. In Figure 5, I look at the

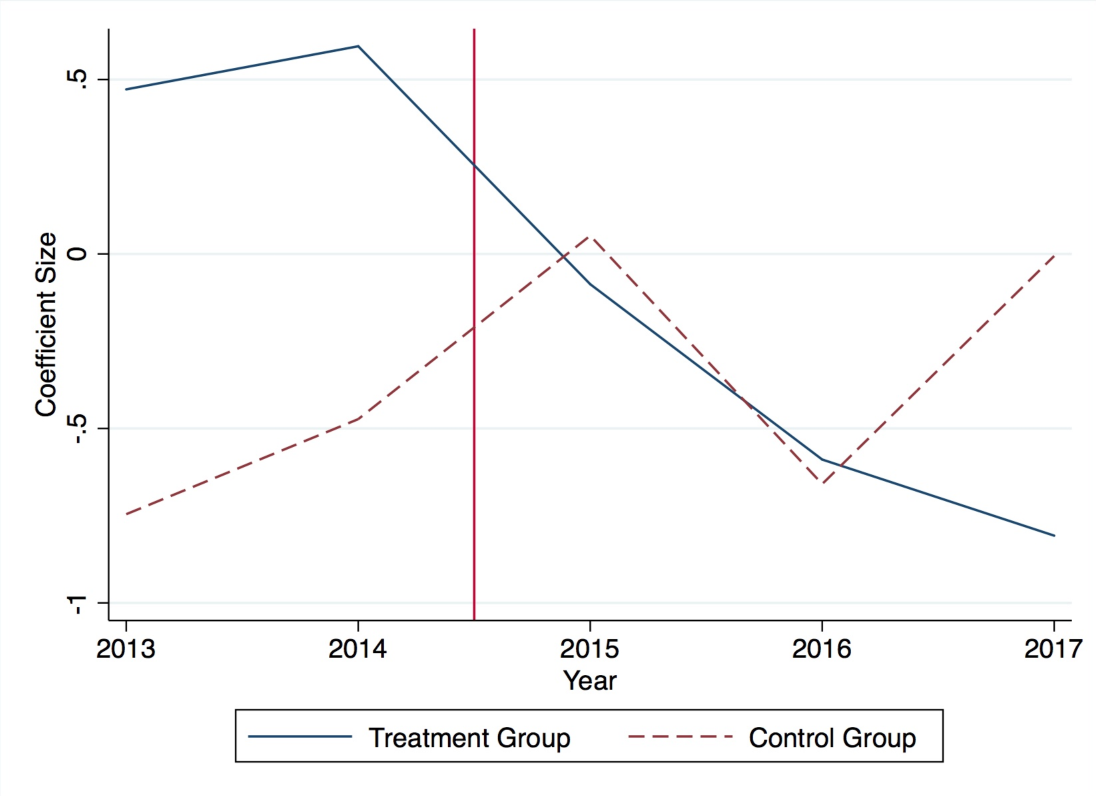
[35]The interactions of these variables with the effect of the reform are not significant at all however

Table 20: Effect of jury characteristics

Dependent variable:	Ranked first		Rank	
Estimator	C. Logit	C. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)
Share of women in committee	-0.993		-0.362	
	(1.04)		(0.62)	
H-index of the president	0.079	0.025	0.014	-0.015
	(0.14)	(0.15)	(0.09)	(0.10)
Age of the Jury President	0.006	0.002	0.003	0.019*
	(0.01)	(0.02)	(0.01)	(0.01)
Mean h-index of the	0.155	0.078	-0.073	-0.171
women in the committee	(0.25)	(0.25)	(0.16)	(0.16)
Mean h-index of the	-0.101	-0.060	0.205	0.312*
men in the committee	(0.26)	(0.28)	(0.17)	(0.19)
Standardised age-adjusted H-index	0.251***	0.252***	0.219***	0.218***
	(0.07)	(0.07)	(0.05)	(0.05)
Share of female candidates	-0.520	-0.466	-0.523	-0.505
	(0.73)	(0.78)	(0.40)	(0.40)
Post Reform · Treatment Group		-1.526***		-0.875***
		(0.57)		(0.31)
Post-reform		0.799**		0.321
		(0.38)		(0.21)
Treatment Group		0.831*		0.383*
		(0.45)		(0.23)
Female president		-0.426		-0.038
		(0.33)		(0.17)
Candidate is a woman		-0.596		-1.168*
		(1.11)		(0.61)
# observations	1332	1332	1332	1332

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

Figure 3: Coefficients from a rank-ordered logit, controlling for publications



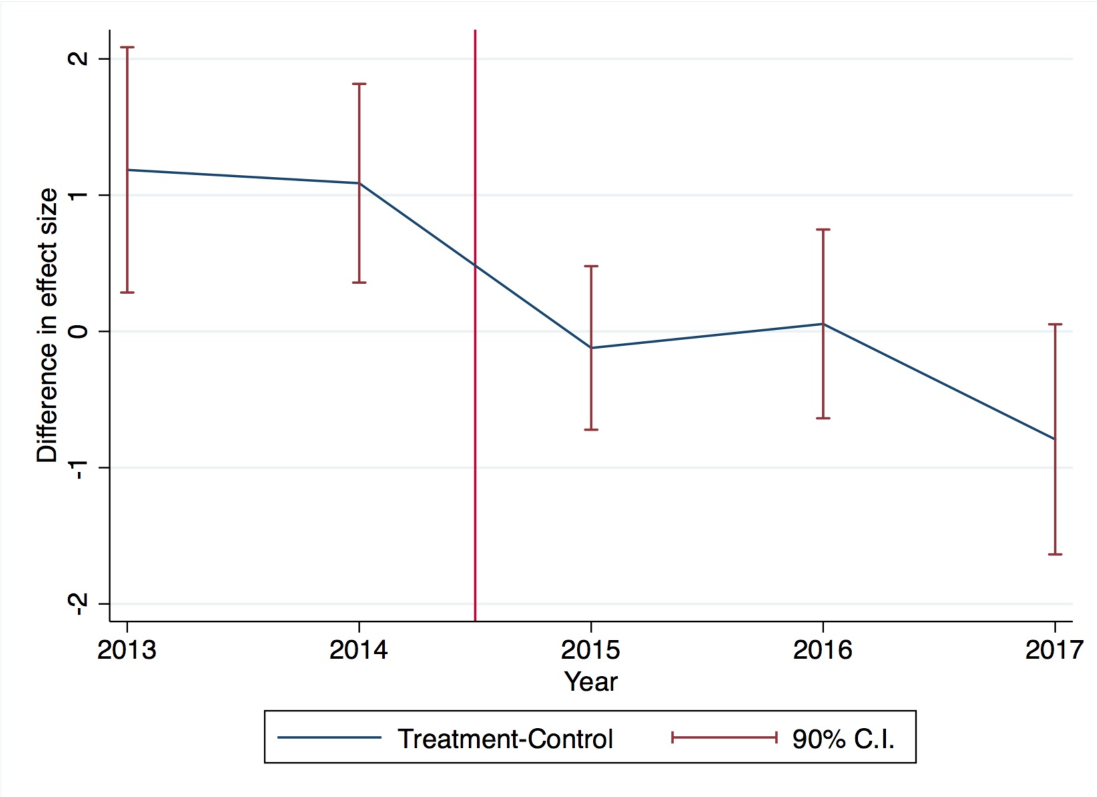
effect of the reform on the ranks of women by discipline[36], with disciplines sorted from left to right according to their pre-reform average share of female jurors. The figure bears out what we see in the regression: disciplines that had few women jurors, and thus were more affected by the quota seem to be less favourable to women after the reform, while disciplines that were already respecting the quota (such as Psychology and Languages), seem to have been positively affected by the reform.

One issue in interpreting these results is that we cannot identify what is driving these worse outcomes, especially since most of the jury level characteristics seem to have little to no effect. One possibility is that women in positions of authority have opposite-gender preferences, a phenomenon that has already been analysed in the sociology literature under the expression "Queen Bee" syndrome.[37] An alternative to the Queen Bee syndrome is that women are penalised when they decide to promote women, as discussed in Johnson and David (2016). The effect that we see is then due to women internalising the retribution they could face from male colleagues if they promote women, rather than opposite-gender preferences. Akerlof and Kranton (2000) postulate instead that men discriminate against

[36]For clarity, smaller fields have been dropped from this figure

[37]see e.g Staines, Tavris, and Jayaratne (1974).

Figure 4: Difference in the effect of gender on ranks between treatment and control group



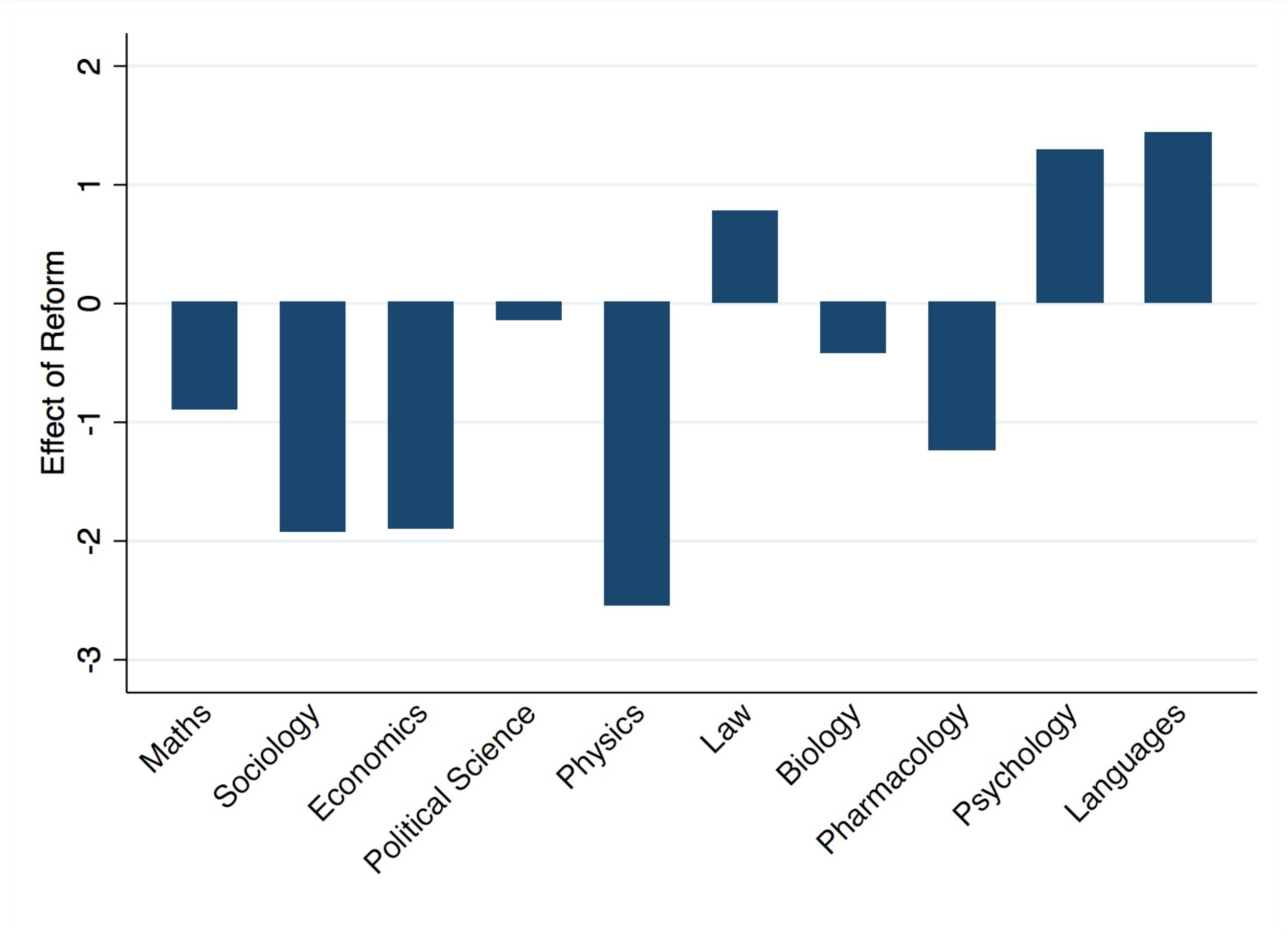
women when their identities are threatened. It is possible that this is the case in our setting, since having more women in what were traditionally masculine settings could be seen as a threat to the masculinity of the jurors. The interpretation of our effects in this case would not be of opposite-gender bias in recruiting, but of own-gender bias from men that appears only when the gender identities of jurors are threatened. Our results are consistent with all three hypotheses.

However, an element that leans towards the third hypothesis is presented in Table 21. In these tables I estimate the following specification:

$$\mathrm{DV}_{ij} = \theta R_j T_j g_i g_j + \alpha R_j T_j g_i + \gamma T_j g_i + \lambda T_j g_i g_j + \delta R_j g_i + \kappa R_j g_i g_j + \beta_1 X_i + \beta_2 g_i Y_j + \epsilon_{ij}$$

where g_j is a dummy variable equal to 1 if the jury president is female. Our parameters of interest here are α and θ, the effect of the reform depending on the gender of the jury president. There is a significant difference between the effect of the reform for committees with male presidents and those with female presidents in the treatment group. The

Figure 5: Effect of the reform on the ranks of women by discipline. Disciplines most affected by the reform are ordered from left to right.



negative effect of the reform is entirely driven by committees with male presidents, for both the conditional and rank-ordered logit specifications. This result is consistent with the idea that men change their behaviour as a result of the reform. [38]

Irrespective of the precise channel that causes this effect, the results presented in this paper strongly suggest that gender quotas and other coercive measures may have unintended effects. In this specific case, increasing the proportion of female jurors has led to a negative effect on the ranking of women by committees.

[38]We can also revisit the effect of the reform on first-ranked candidates in the probit specification. Interacting the effect of the reform with the gender of the jury president leads to similar results on hiring, though the results are barely significant at the 10% level. These results are presented in Table 31 of Appendix F

Table 21: Effect of the reform on women by gender of the jury president

Dependent variable:	Ranked first		Rank	
Estimator:	C. Logit	C. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)	(4)
Post Reform · Treatment Group	-2.217***	-2.628***	-1.147***	-1.395***
· Male President	(0.66)	(0.75)	(0.37)	(0.47)
Post Reform · Treatment Group	2.471**	3.076**	1.189*	1.089
· Female President	(1.25)	(1.33)	(0.64)	(0.70)
Treatment Group· Female president	-1.383	-1.570*	-0.793*	-0.880**
	(0.93)	(0.94)	(0.44)	(0.44)
Treatment Group· Male president	1.200**	0.754	0.579**	0.124
	(0.49)	(0.83)	(0.27)	(0.40)
Control Group·Post-reform	1.170**	1.441***	0.448*	0.381
· Male president	(0.47)	(0.53)	(0.25)	(0.29)
Control Group·Post-reform	-1.247	-1.530*	-0.481	-0.455
· Female president	(0.77)	(0.90)	(0.43)	(0.49)
	-0.741	-1.376	-0.220	-0.184
	(0.56)	(0.99)	(0.26)	(0.46)
Candidate is a woman· Female president	0.218	0.203	0.266	0.153
	(0.58)	(0.67)	(0.30)	(0.34)
Share of female candidates	-0.324	0.828	-0.429	-0.207
	(0.76)	(0.85)	(0.40)	(0.44)
Standardised age-adjusted H-index		0.370***		0.286***
		(0.08)		(0.05)
PhD supervisor in the committee		0.465*		0.365**
		(0.27)		(0.16)
PhD from the same institute		0.593***		0.317***
		(0.18)		(0.10)
Professorship		0.914***		0.368**
		(0.31)		(0.18)
Discipline Fixed Effects	No	Yes	No	Yes
# observations	1357	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. Contests where only one candidate was ranked, or where only applicants of a specific gender applied are dropped from the analysis.

7 Conclusion

In this article, I have showed that imposing gender quotas in recruitment committees does not help the recruitment of women, and may actually harm their careers. It is unclear whether this effect is driven by women being less likely to favour women in hiring, or through men voting against women as a reaction to the quota, though there is some evidence in favour of the second channel. The results do not preclude there being discrimination against women in hiring,[39] but they do suggest that there may not be an easy solution to the problem of under-representation of women in academia.

Gender quotas were successful in increasing the proportion of women in the committees, less so in ensuring that these women were in fact in positions of power since the proportion of female committee presidents did not change following the reform. I find no evidence that the new jury members are less qualified, and no evidence either that the reform encouraged more women to apply to professorial positions, at least in the short term. Most of the new jury members are internal members, which suggests that relative to their male colleagues, the administrative burden of women in the departments increases. A global evaluation of the impact of the reform depends on the weights one attaches to each of these aspects.

References

Abrevaya, J., and D. S. Hamermesh (2012): "Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?," *The Review of Economics and Statistics*, 94(1), 202–207.

Ahern, K., and A. K. Dittmar (2012): "The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation," *The Quarterly Journal of Economics*, 127(1), 137–197.

Akerlof, G. A., and R. E. Kranton (2000): "Economics and identity," *The Quarterly Journal of Economics*, 115(3), 715–753.

Antecol, H., K. Bedard, and J. Stearns (2018): "Equal but Inequitable: Who Benefits from Gender-Neutral Tenure Clock Stopping Policies?," *American Economic Review*, 108(9), 2420–41.

Azmat, G. (2014): "Gender diversity in teams," *IZA World of Labor*.

Bagues, M., M. Sylos-Labini, and N. Zinovyeva (2017): "Does the gender composition of scientific committees matter?," *American Economic Review*, 107(4), 1207–38.

Bagues, M. F., and B. Esteve-Volart (2010): "Can Gender Parity Break the Glass Ceiling?

[39] It seems in fact quite likely that this is the case. See Table 6 for instance, where the coefficient in column 1 for the effect of being a woman on rank is negative.

Evidence from a Repeated Randomized Experiment," *Review of Economic Studies*, 77(4), 1301–1328.

Bagues, M. F., and N. Zinovyeva (2015): "The Role of Connections in Academic Promotions," *American Economic Journal: Applied Economics*, 124(5).

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009): "Powerful women: does exposure reduce bias?," *The Quarterly Journal of Economics*, 124(4), 1497–1540.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012): "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India," *Science*, 335(6068), 582–586.

Beggs, S., S. Cardell, and J. Hausman (1981): "Assessing the potential demand for electric cars," *Journal of Econometrics*, 17(1), 1–19.

Bertrand, M. (2011): "New perspectives on gender," in *Handbook of Labor Economics*, vol. 4, pp. 1543–1590. Elsevier.

Bertrand, M., S. E. Black, S. Jensen, and A. Lleras-Muney (2018): "Breaking the glass ceiling? The effect of board quotas on female labor market outcomes in Norway," *Review of Economic Studies*.

Boring, A. (2017): "Gender biases in student evaluations of teaching," *Journal of Public Economics*, 145(C), 27–41.

Bosquet, C., P.-P. Combes, and C. Garcia-Peñalosa (2018): "Gender and Promotions: Evidence from Academic Economists in France," *Scandinavian Journal of Economics*.

Breda, T., and M. Hillion (2016): "Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France," *Science*, 353(6298), 474–478.

Breda, T., and S. T. Ly (2015): "Professors in Core Science Fields Are Not Always Biased against Women: Evidence from France," *American Economic Journal: Applied Economics*, 7(4), 53–75.

Broder, I. E. (1993): "Review of NSF Economics Proposals: Gender and Institutional Patterns," *The American Economic Review*, 83(4), 964–970.

Buser, T., M. Niederle, and H. Oosterbeek (2014): "Gender, competitiveness, and career choices," *The Quarterly Journal of Economics*, 129(3), 1409–1447.

Casas-Arce, P., and A. Saiz (2015): "Women and power: unpopular, unwilling, or held back?," *Journal of Political Economy*, 123(3), 641–669.

Chenu, A., and O. Martin (2016): "Le plafond de verre chez les enseignants-chercheurs en sociologie et démographie," *Travail, genre et sociétés*, (2), 133–156.

Colussi, T. (2018): "Social Ties in Academia: A Friend Is a Treasure," *The Review of Economics and Statistics*, 100(1), 45–50.

Combes, P.-P., L. Linnemer, and M. Visser (2008): "Publish or peer-rich? The role of skills and networks in hiring economics professors," *Labour Economics*, 15(3), 423–441.

Duflo, E. (2000): "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," Discussion paper, National Bureau of Economic Research.

Edo, A., N. Jacquemet, and C. Yannelis (2015): "Language skills and homophilous hiring discrimination: Evidence from gender and racially differentiated applications," *Review of Economics of the Household*, pp. 1–28.

Ellemers, N., H. Van den Heuvel, D. de Gilder, A. Maass, and A. Bonvini (2004): "The underrepresentation of women in science: Differential commitment or the queen bee syndrome?," *British Journal of Social Psychology*, 43(3), 315–338.

Feld, J., N. Salamanca, and D. S. Hamermesh (2016): "Endophilia or exophobia: beyond discrimination," *The Economic Journal*, 126(594), 1503–1527.

Geuna, A., and S. Shibayama (2015): "Moving out of academic research: why do scientists stop doing research?," in *Global mobility of research scientists*, pp. 271–303. Elsevier.

Hardle, W., and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits," *The Annals of Statistics*, pp. 1926–1947.

Hausman, J. A., and P. A. Ruud (1987): "Specifying and testing econometric models for rank-ordered data," *Journal of Econometrics*, 34(1-2), 83–104.

Ianni, G. (2010): "Scholar H-index Calculator," .

Johnson, S. K., and H. R. David (2016): "Women and minorities are penalized for promoting diversity," *Harvard Business Review. Retrieved from https://hbr. org/2016/03/women-andminorities-are-penalized-for-promoting-diversity.*

Li, D. (2017): "Expertise versus Bias in Evaluation: Evidence from the NIH," *American Economic Journal: Applied Economics*, 9(2), 60–92.

Lincoln, A. E., S. Pincus, J. B. Koster, and P. S. Leboy (2012): "The Matilda Effect in science: Awards and prizes in the US, 1990s and 2000s," *Social Studies of Science*, 42(2), 307–320.

Nimmesgern, H. (2016): "Why Are Women Underrepresented in STEM Fields?," *Chemistry – A European Journal*, 22(11), 3529–3530.

Paola, M. D., and V. Scoppa (2015): "Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia," *Economica*, 82(325), 162–188.

Rebérioux, A., and G. Roudaut (2017): "Gender Quota and Inequalities inside the Boardroom," .

Robinson, P. M. (1988): "Root- N-Consistent Semiparametric Regression," *Econometrica*, 56(4), 931–954.

Staiger, D., and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, vol.65(no.3), 557–586.

Staines, G., C. Tavris, and T. E. Jayaratne (1974): "The queen bee syndrome," *Psychology Today*, 7(8), 55–60.

Thurstone, L. L. (1927): "A Law of Comparative Judgement," *Psychological Review*, 34, 278–286.

Verardi, V., and N. Debarsy (2012): "SEMIPAR: Stata module to compute Robinson's (1988) semiparametric regression estimator," Statistical Software Components, Boston College Department of Economics.

Vernos, I. (2013): "Research management: Quotas are questionable," *Nature*, 495(7439), 39–39.

Williams, W. M., and S. J. Ceci (2015): "National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track," *Proceedings of the National Academy of Sciences*, 112(17), 5360–5365.

Zeltzer, D. (2015): "Gender Homophily in Referral Networks: Consequences for the Medicare Physician Pay Gap," *Working Paper*.

A Semiparametric estimation

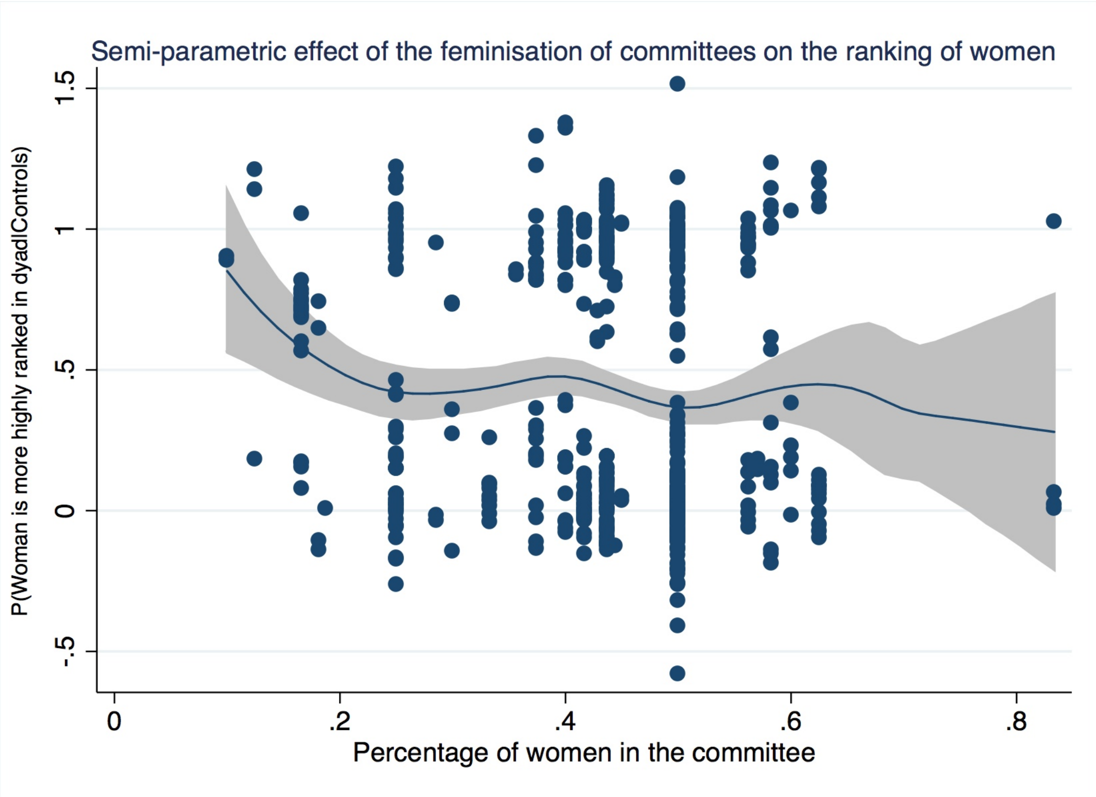
One hypothesis from Paola and Scoppa (2015) was that the effect of the feminisation of committees on the likelihood of being promoted was non-linear. In their paper, they found that there was a positive effect on the probability of a woman being promoted when committees moved from having no women to having 1 woman, and no further effect thereafter. Although we do not observe committees with no women in our dataset, we can check whether assuming a linear effect of feminisation of the committee is correct.

Consider the following partially linear model,

$$W_{ij} = X\beta + m(Gg_j) + \epsilon_{ij}$$

where W_{ij} is a binary variable, equal to 1 if a women is better ranked than a man for a specific dyad, X, β and G are defined in section 3 and m is a function that we wish to estimate. Using Robinson (1988)'s double residual estimator, we can first estimate β consistently, then recover m(Gg_j) non-parametrically. As can be seen below, from a semi-parametric regression that controls for the h-index of candidates, the effect of the feminisation of committees on the ranking of women does not seem to be non-linear.

Figure 6: Semi-parametric estimation of the relationship of interest



Using Hardle and Mammen (1993)'s test, that relies on square differences between a linear estimator and the estimated non-parametric function, we cannot reject the null hypothesis that the linear form and the non-parametric estimates are not different.

B Differences between the Random Utility Model and the Rank-Ordered Logit

Are the two methods presented in the article equivalent? Consider the following data generating process. I generate a jury utility variable U_{ij} such that:

$$U_{ij} = \beta q_{ij} + \mu g_i g_j + \epsilon_{ij}$$

where g_i and g_j are binomial variables, q_{ij} is normally distributed with mean 5 and standard deviation of 1. Using the values from the article, I use 0.05 for β and -0.6 for μ.

From this utility variable I then simulate artificial committees in the following way: from a 1000 observations of U_{ij} I create 100 committees with 10 candidates in each committee. 60 committees have a female jury president. Each committee then creates a ranking r_{ij} of each candidate based on the value of U_{ij}.

This simulation allows me to answer two different questions: do both methods have the right power, and can both methods accurately recover the μ/β ratio?

Table 22: Power of estimation methods

Estimator	RO Logit	OLS	Probit	Logit
Normal errors				
Mean	-0.03	-0.00	-0.00	0.00
SD	0.11	0.04	0.04	0.04
# of t-statistics $\leq$|1.96| (%)	947 (95%)	941(94%)	941(94%)	941(94%)
Type I Extreme Value				
Mean	0.00	0.00	0.00	0.00
SD	0.11	0.04	0.04	0.04
# of t-statistics $\geq$|1.96| (%)	943 (94%)	960(96%)	960(96%)	959(96%)

Results from 1000 simulations. True effect is 0. For the right amount of power, tests should reject the null hypothesis only 5% of the time. Results from the probit and logit estimations are marginal effects.

Table 23: Estimates from simulations

Estimator	RO Logit	OLS	Probit	Logit
Gender bias				
Mean	-0.61	-0.15	-0.38	-0.60
SD	0.11	0.04	0.10	0.17
Effect of quality				
Mean	0.05	0.01	0.03	0.05
SD	0.04	0.01	0.03	0.04
Ratio				
Median	-0.086	-0.086	-0.85	-0.085
SD	0.07	0.10	0.10	0.10

Means and standard deviation of estimators of μ and β from 1000 simulations. Error term is distributed type I EV. Results are similar with normally distributed errors. Results from the probit and logit estimations are full effects and not marginal. The marginal effects are similar in size to the OLS results.

Let us consider the first question. The right standard errors would give us a 5% probability of having a t-statistic above the absolute value of 1.96, if the true effect was in fact 0. I simulate the DGP described above 500 times, with a true effect of gender bias, $\mu =$ to 0. I present results both with ϵ_{ij} either normally distributed or type I extreme value in Table 22. In all cases, errors are clustered at the contest level, since candidates evaluated by the same jury are compared multiple times, generating correlation between errors.

The methods that use the random utility approach are very similar in terms of power and estimates. They seem to over-reject when errors are distributed Extreme Value, and under-reject when errors are distributed normally. However, the power estimates are within 1% of the valid rejection rates. The rank-ordered logit has a wider standard deviation of estimates, but is closer to the valid rejection rates.

In Table 23, we can see the mean values of the estimators for the parameters presented above. I present the median value for the μ/β ratio, since the mean is affected by outliers, i.e. when β is estimated to be very close to 0. The mean point estimates are similar for the rank-ordered logit, probit and OLS. The results are similar no matter the method considered.

This is the case when the values of the parameters are small. However, with larger values of β and μ, the LPM suffers from bias, as can be seen in Table 24. This well known result stems from the predicted values of Y, $X\beta + \mu G_{ij}$, being outside the range of possible values that can be taken on by a binary variable, i.e. 0-1.

Table 24: Estimates from simulations: large parameters

Estimator	RO. Logit	OLS	Probit	Logit
Gender bias				
Mean	-6.01	-0.40	-3.33	-6.04
SD	0.30	0.02	0.20	0.36
Effect of quality				
Mean	3.00	0.22	1.66	3.01
SD	0.13	0.01	0.08	0.14
Ratio				
Median	-0.50	-0.55	-0.50	-0.50
SD	0.02	0.02	0.02	0.02

Means and standard deviation of estimators of μ and β from 1000 simulations. Error term is distributed type I EV. Results are similar with normally distributed errors. Results from the probit and logit estimations are full effects and not marginal. The marginal effects are similar in size to the OLS results.

C Probit and heteroscedasticity

In this section I present results using an instrument variable probit. Although the endogenous variable, the share of women in the jury, is continuous, the instrument considered is binary which may introduce heteroscedasticity. Under heteroscedasticity, the results from probit regressions are biased, which is why a LPM is used in the main text. The marginal effects are significant, but slightly smaller in magnitude compared to those in the main text.

Table 25: IV Probit results

Dependent variable:	Rank		
Estimator:	Probit	Probit	Probit
Marginal Effects	(1)	(2)	(3)
Share of women in the committee	-1.593***	-1.564***	-1.535***
	(0.41)	(0.41)	(0.44)
Female president	0.022	0.018	0.017
	(0.05)	(0.05)	(0.05)
Treatment Group	-0.150***	-0.146***	-0.147***
	(0.05)	(0.05)	(0.05)
Reform	0.067	0.062	0.062
	(0.04)	(0.04)	(0.04)
Standardised H-index difference		0.065***	0.066***
		(0.01)	(0.02)
PhD from the same institute		0.040	0.042
		(0.03)	(0.03)
Candidate(W)'s PhD advisor is in the committee		0.051	0.053
		(0.05)	(0.05)
Share of female candidates			0.001
			(0.14)
Professorship			0.024
			(0.05)
# observations	1086	1086	1086

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Errors are clustered at the contest level.

Table 26: Probit results, Difference-in-Difference

Dependent variable:	Rank		
Estimator:	Probit	Probit	Probit
Marginal effects	(1)	(2)	(3)
Post Reform*Treatment Group	-0.285***	-0.280***	-0.289***
	(0.09)	(0.09)	(0.08)
Female president	-0.031	-0.030	-0.030
	(0.05)	(0.05)	(0.05)
Reform	0.135***	0.133***	0.139***
	(0.05)	(0.05)	(0.05)
Treatment Group	0.155**	0.154**	0.117
	(0.07)	(0.07)	(0.07)
H-index difference		0.014***	0.014***
		(0.00)	(0.00)
PhD from the same institute		0.038	0.045
		(0.03)	(0.03)
Candidate(W)'s PhD advisor is in the committee		0.067	0.066
		(0.05)	(0.05)
Share of female candidates			-0.219
			(0.13)
# observations	1086	1086	1086

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Errors are clustered at the contest level.

D Difference in difference

I present results in Tables 26 and 27 that are similar to the graphical interpretation in the text. Instead of instrumenting the share of women in the jury, I rely on a difference in difference analysis, comparing variation in the probability of women being hired before and after the reform in treatment and control group. Being in the treatment group after the reform significantly decreases the probability of a woman being hired, even when controlling for H-index differences and connections. The point estimates found using probit and OLS are extremely similar.

E Full effect of the reform

In this section I look at the direct effect of the reform without considering heterogeneity in treatment effects across disciplines, using only a dummy variable for the effect of the reform. With this methodology, the net effect of the reform seems to be close to 0. The estimate is insignificant, and the point estimate of the effect is 4 times smaller than that of a standard deviation in age-discounted h-index.

F Additional tables

To see whether publications could explain gender difference in hiring, I look at gender differences in h-indices and Citations in Table 29, regressing 4 measures of publication on a set of fixed effects, which include whether the post is for a full professorship, field fixed effects, and interactions between field and professorship. To account for the large field differences in citations, I standardise the h-index measures by field in the main parts of the estimation. This variable is included in column 4. Female candidates have lower measures of publication quality no matter the measure chosen, though the effect on citations is not significant. It is therefore crucial to control for differences in publications in the estimation.

In Table 30, I look at the correlation between the share of women in committees and using the random utility model. In this specification, there is no significant effect of either the gender of the jury president, or the percentage of women in the committee.

Table 27: OLS results, Difference-in-Difference

Dependent variable:	Rank		
Estimator:	OLS	OLS	OLS
	(1)	(2)	(3)
Post Reform*Treatment Group	-0.301***	-0.298***	-0.304***
	(0.09)	(0.09)	(0.09)
Reform	0.150***	0.149***	0.154***
	(0.05)	(0.05)	(0.05)
Treatment Group	0.174**	0.175**	0.129*
	(0.07)	(0.07)	(0.07)
H-index difference		0.010***	0.010***
		(0.00)	(0.00)
PhD from the same institute		0.037	0.044
		(0.03)	(0.03)
Candidate(W)'s PhD advisor is in the committee		0.064	0.064
		(0.05)	(0.05)
Share of female candidates			-0.261**
			(0.13)
Discipline Fixed Effects	No	No	No
# observations	1093	1093	1093

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Errors are clustered at the contest level.

In Table 31, I present an additional specification, using a difference-in-difference approach to estimate the effect of the reform on hiring, rather than an IV as in the main section of the paper. The results are very similar, but this apporach allows me to split the effect of the reform by gender of the jury president. Though these results are barely significant, they seem to point in the same direction as those presented in Section 6 of the paper; Juries with male presidents in the treated groups hire less women after the reform.

In Table 32, I use the same specification as in Table 10, but include other measures of publications that could be more relevant than the h-index. I include variables with age- and author-discounted h-indices and citations. The variable that seems most significant in this case the age-discounted h-index, which is why I choose this variable in the main specifications of the article as a control for publications. This table include the h-index of the jurors, and the mean age of the jury members, for which I have data for a subset of the data only. Including these other citation measures makes no difference to the effect of the reform. There is no effect of the mean jury h-index, however older committees seem to view women more positively.

In Table 33, the sample is split into two different categories: STEM (Chemistry, Engineering, Pharmacology, Physics and Maths) and non-STEM, since a lot of the debate around the under-representation of women is about STEM fields (e.g. Nimmesgern (2016)). Buser, Niederle, and Oosterbeek (2014) find that men are more likely to major in STEM fields than women with equivalent grades. There is also some evidence that women in France receive better evaluations in fields where they are under-represented.[40] I find that the effect of having more women in the jury is negative in both STEM and non-STEM, but the point estimate in STEM fields is higher. This could simply be picking up part of the effect of the reform however. The difference between the coefficients is not significant however.

In Table 34, I look at how the share of women among ranked candidates evolves with the reform. I use the same methodology as in Table 15 Here there seems to be a slight concern since women are more likely to be ranked after the reform in the treatment group, though the difference between treatment and control group is not significant once we limit the analysis to two years either side of the reform.

In Table 35, I try to see whether dropping lower ranked candidates can help mitigate these concerns. This is the sample from Table 17. With this specification, we see that the point estimate

[40] e.g.Breda and Ly (2015) and Breda and Hillion (2016).

Table 28: Global effect of the reform

Dependent variable:	Rank		
Estimator:	RO. Logit	RO. Logit	RO. Logit
	(1)	(2)	(3)
Post-reform	-0.048	-0.063	-0.099
	(0.15)	(0.15)	(0.19)
Candidate is a woman	-0.213**	-0.031	0.047
	(0.11)	(0.22)	(0.32)
Standardised age-adjusted H-index		0.275***	0.280***
		(0.04)	(0.05)
Share of female candidates		-0.302	-0.400
		(0.39)	(0.43)
PhD supervisor in the committee			0.350**
			(0.15)
PhD from the same institute			0.296***
			(0.10)
Discipline fixed effects	No	No	Yes
# observations	1357	1357	1357

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. The h-index variable can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men.

Table 29: Gender differences in h-indices

Dependent variable:	H-index	Age-adjusted H-index	Citations	Standardised age-adjusted H-index
Estimator:	OLS	OLS	OLS	OLS
Candidate is a woman	-1.152***	-0.186	-49.856	-0.131**
	(0.29)	(0.12)	(54.42)	(0.05)
Fixed effects	Yes	Yes	Yes	Yes
# observations	1548	1548	1548	1548

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Fixed effects include University, whether the job is for a professorship, field and field · professorship interaction terms.

for the change between treatment and control group following the reform is much smaller than in the regular sample, especially once we consider only the two years either side of the reform. Since this sample is actually the one that gives the highest point estimates in the paper, it seems unlikely that changes in selection into the ranking pool drives the results.

In Table 36, I test for pre-trends between the treatment and the control group, with the three methods used in the main part of the article (conditional logit, rank-ordered logit, probit on dyads). I find no significant trend, though the point estimates are high.

Table 30: Correlation between the gender of jurors and the probability that a woman is better ranked within a dyad

Dependent variable:	Probability that W is more highly ranked			
Estimator:	Probit	Probit	Probit	Probit
Marginal effects	(1)	(2)	(3)	(4)
Female president	-0.020	-0.021	-0.034	-0.031
	(0.05)	(0.04)	(0.05)	(0.05)
Share of women in the committee	-0.136	-0.086	-0.141	-0.150
	(0.20)	(0.19)	(0.21)	(0.21)
Standardised H-index difference		0.084***	0.087***	0.087***
		(0.01)	(0.01)	(0.01)
Share of female candidates			-0.091	-0.104
			(0.18)	(0.19)
Professorship			0.082*	0.086*
			(0.05)	(0.05)
Candidate(W)'s PhD advisor is in the committee				0.062
				(0.05)
PhD from the same institute				0.044
				(0.03)
Discipline fixed effects	No	No	Yes	Yes
University fixed effects	No	No	Yes	Yes
# observations	1086	1086	1086	1086

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Standard errors are clustered at the jury level. Variables on the PhD's of candidates are coded as follows: If both or neither candidates have an advisor in the committee, then the variable is equal to 0. If a candidate has an advisor in the committee and not the other, then the variable is coded either as 1 or -1, depending on the gender of the connected candidate.

Table 31: Probit on gender of first-ranked candidate

Dependent variable:	Gender of first-ranked candidate			
Estimator:	Probit	Probit	Probit	Probit
	(1)	(2)	(3)	(4)
Treatment Group	-0.437**	0.174	0.359	0.240
	(0.19)	(0.22)	(0.27)	(0.29)
Reform	0.246	0.232	0.322	0.338
	(0.17)	(0.19)	(0.22)	(0.22)
Post Reform·Treatment Group	-0.198	-0.340	-0.560*	-0.562
	(0.25)	(0.28)	(0.34)	(0.35)
Female president		-0.062		
		(0.15)		
Share of female candidates		3.355***	3.428***	2.624***
		(0.29)	(0.30)	(0.45)
Post Reform·Treatment Group			0.557	0.584
·Female President			(0.58)	(0.59)
Post Reform·Female President			-0.168	-0.180
			(0.24)	(0.24)
Treatment group·Female President			-0.304	-0.356
			(0.40)	(0.41)
University fixed effects	No	No	Yes	Yes
# observations	455	455	455	359

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. This regression considers first-ranked candidates only. Contests with applicants from a specific gender only are dropped in column four.

Table 32: Rank-Ordered Logit using the quota: Other publication measures

Dependent variable:	Rank		
Estimator:	Logit	Logit	Logit
	(1)	(2)	(3)
Post Reform · Treatment Group	-0.844***	-0.838***	-1.069***
	(0.31)	(0.30)	(0.40)
Post-reform	0.309	0.307	0.232
	(0.20)	(0.20)	(0.23)
Treatment Group	0.433**	0.382*	-0.123
	(0.22)	(0.22)	(0.30)
Female president	-0.090	-0.071	-0.141
	(0.17)	(0.16)	(0.17)
Candidate is a woman	-0.320**	-0.127	-0.071
	(0.15)	(0.25)	(0.40)
H-index of the candidate	-0.014	-0.014	-0.015
	(0.02)	(0.02)	(0.03)
Age-discounted H-index	0.149***	0.149***	0.146***
	(0.05)	(0.05)	(0.05)
Co-author discounted	0.038	0.038	0.046
H-index of the candidate	(0.04)	(0.04)	(0.04)
Citations of the candidate	-0.000	-0.000	0.000
	(0.00)	(0.00)	(0.00)
Co-author discounted	0.000	0.000	-0.000
citations of the candidate	(0.00)	(0.00)	(0.00)
Age-Discounted citations	-0.001	-0.001	-0.001
of the candidate	(0.00)	(0.00)	(0.00)
Share of female candidates		-0.392	-0.188
		(0.41)	(0.43)
Professorship			0.401**
			(0.20)
PhD supervisor in the committee			0.387**
			(0.16)
PhD from the same institute			0.296***
			(0.10)
Discipline Fixed Effects	No	No	Yes
# observations	1355	1355	1355

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. The h-index variable and other citation measures can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men.

Table 33: Correlation between the gender of jurors and the probability that a woman is better ranked within a dyad. Split by subject

Dependent variable:	Probability that W is more highly ranked			
Marginal effects	(1)	(2)	(3)	(4)
Female president	0.005	-0.011	0.041	-0.026
	(0.08)	(0.05)	(0.08)	(0.06)
Share of women in the committee	-0.563	0.060	-0.575	-0.060
	(0.41)	(0.22)	(0.36)	(0.26)
Share of female candidates	0.295	-0.062	0.170	-0.313
	(0.29)	(0.17)	(0.35)	(0.20)
H-index difference			0.018***	0.014***
			(0.01)	(0.00)
PhD supervisor in the committee			-0.070	0.135**
			(0.08)	(0.06)
PhD from the same institute			0.059	0.031
			(0.06)	(0.04)
Discipline	STEM	Social Science	STEM	Social Science
Discipline fixed effects	No	No	Yes	Yes
University fixed effects	No	No	Yes	Yes
# observations	397	689	397	689

$*p < 0.10, **p < 0.05, ***p < 0.01$. Dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Standard errors are clustered at the jury level

Table 34: Effect of the reform on the share of women among ranked candidates

Dependent variable:	Gender composition of ranked candidates					
Marginal effects	(1)	(2)	(3)	(4)	(5)	(6)
Reform	0.002	0.110**	-0.005	-0.011	0.072	-0.018
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)
Professorship	-0.178***	-0.104**	-0.145***	-0.174***	-0.058	-0.127***
	(0.05)	(0.04)	(0.03)	(0.06)	(0.05)	(0.04)
Treatment Group			-0.055			0.005
			(0.07)			(0.10)
Treatment Group * Reform			0.109*			0.077
			(0.06)			(0.07)
Group	Control Group	Treatment Group	Both	Control Group	Treatment Group	Both
Discipline Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# observations	235	220	455	166	130	296

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The last 3 columns are restrained to 2013-2016.

Table 35: Effect of the reform on the share of women among ranked candidates- Top 3 Ranks only

Dependent variable:	Gender composition of ranked candidates					
	(1)	(2)	(3)	(4)	(5)	(6)
Reform	0.025	0.111**	0.019	0.020	0.050	0.015
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)
Professorship	-0.162***	-0.098**	-0.133***	-0.141**	-0.049	-0.104**
	(0.05)	(0.05)	(0.03)	(0.06)	(0.06)	(0.04)
Treatment Group			-0.091			0.032
			(0.07)			(0.11)
Treatment Group*Reform			0.085			0.028
			(0.07)			(0.08)
Group	Control Group	Treatment Group	Both	Control Group	Treatment Group	Both
Discipline Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
# observations	235	220	455	166	130	296

$*p < 0.10, **p < 0.05, ***p < 0.01$. The gender variable is coded as 1 for a woman and 0 for a man. Only the top 3 candidates are considered. Treatment variable here is based on mean pre-reform gender parity at the field*university level. The last 3 columns are restrained to 2013-2016.

Table 36: Testing for pre-trends

	Rank	First Ranked	Winner of Dyad
Estimator	R.O. Logit	C. Logit	Probit
Treatment Group	0.449	-0.594	0.324
	(1.40)	(3.06)	(0.35)
Trend in Treatment group	0.504	0.586	0.060
	(0.32)	(0.65)	(0.07)
Trend in Control group	-0.193	0.090	0.006
	(0.26)	(0.50)	(0.05)
Standardised age-adjusted H-index	0.269***	0.360***	0.062***
	(0.06)	(0.12)	(0.02)
Share of female candidates	0.540	1.400	0.275
	(0.65)	(1.55)	(0.26)
Candidate is a woman	-2.428*	-2.989	.
	(1.34)	(2.79)	.
PhD supervisor in the committee	0.650**	0.730*	0.196**
	(0.27)	(0.43)	(0.08)
PhD from the same institute	0.094	0.456	0.021
	(0.19)	(0.30)	(0.05)
Professorship	0.610**	1.174**	0.206***
	(0.29)	(0.54)	(0.07)
Discipline Fixed Effect	Yes	Yes	Yes
Time Period	2009-2014	2009-2014	2009-2014
Total	555	555	407

$*p < 0.10, **p < 0.05, ***p < 0.01$. For columns 1 and 2, the gender variable is coded as 1 for a woman and 0 for a man. The h-index variable and other citation measures can be interpreted as the effect of increasing the h-index for all candidates. Other variables must be interpreted as the differential effect of the variable for women with respect to men. For column 3, dyads represent pairs of female-male candidates ranked by the same jury. The gender variable is coded as 1 for a woman and 0 for a man. Standard errors are clustered at the jury level. Variables on the PhD's of candidates are coded as follows: If both or neither candidates have an advisor in the committee, then the variable is equal to 0. If a candidate has an advisor in the committee and not the other, then the variable is coded either as 1 or -1, depending on the gender of the connected candidate.

Université Sorbonne
Paris Cité

Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex). Ce projet est distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR). Il est financé dans le cadre des investissements d'avenir.

(ANR-11-LABX-0091, ANR-11-IDEX-0005-02)

www.sciencespo.fr/liepp

A propos de la publication

Procédure de soumission :
Rédigé par un ou plusieurs chercheurs sur un projet en cours, le *Working paper* vise à susciter la discussion scientifique et à faire progresser la connaissance sur le sujet étudié. Il est destiné à être publié dans des revues à comité de lecture (peer review) et à ce titre répond aux exigences académiques. Les textes proposés peuvent être en français ou en anglais. En début de texte doivent figurer : les auteurs et leur affiliation institutionnelle, un résumé et des mots clefs.

Le manuscrit sera adressé à : liepp@sciencespo.fr

Les opinions exprimées dans les articles ou reproduites dans les analyses n'engagent que leurs auteurs.

Directeur de publication :
Bruno Palier

Comité de rédaction :
Andreana Khristova, Juliette Seban

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61
liepp@sciencespo.fr