Research Similarity and Women in Academia *

Piera Bello, Alessandra Casarico, Debora Nozza

October 17, 2025

Abstract

We investigate the extent to which research similarity between senior and junior researchers is related to promotion in academia and study implications for gender diversity among academic staff. Using data on the universe of job applications for tenure track assistant professor positions in economics in Italy, and applying NLP techniques (i.e., document embeddings) to the abstract of each publication of the scholars in our dataset, we propose a novel measure of research similarity that can capture the closeness in research topics, methodologies or policy relevance between candidates and members of selection committees. We show that the degree of similarity is strongly associated with the probability of winning. Moreover, while there are no gender differences in mean similarity, the maximum similarity with selection committee members is lower for female candidates. This gender gap disappears when similarity is calculated focusing only on female committee members. The results suggest that similarity bias in male-dominated environments may have implications for gender and research diversity.

*Corresponding author: Piera Bello, Department of Economics, University of Bergamo, piera.bello@unibg.it. For useful comments, we thank Germain Gauthier, Salvatore Lattanzio, and participants at the Workshop on Gender Discrimination, University of Bergamo; Workshop on Gender and Institutions, University of Venice, Ca'Foscari; the SPS Conference on Global Challenges; the SABE-IAREP Conference; the Frogee-FREE Network conference on The playing field in Academia in Warsaw; the Workshop for Women in Political Economy, Stockholm School of Economics; the SIEP Conference in Verona; the EALE conference in Bergen; the DAISI Workshop at Bocconi University. We also thank seminar participants at University College Dublin; Economics and Data Science, ETH Zurich; the LSE Department of Social Policy; the Summer School in Computational Social Sciences at the University of Bologna. Laura Lufray, Nina Lurati, Julia Milis, and Isabella Masetto have provided excellent research assistance. Casarico acknowledges that this research is part of the MUSA – Multilayered Urban Sustainability Action – project, funded by the European Union –NextGenerationEU, under the National Recovery and Resilience Plan (NRRP). Bello acknowledges that this research is part of the National Recovery and Resilience Plan (NRRP) project "Break the bias: new data and methods to investigate the presence and impact of gender stereotypes on economic outcomes", funded by the European Union – Next Generation EU, ID MUR: 2022AE5BCC_01, CUP: F53D23003020006 (MUR). Usual disclaimers apply. Declarations of interest: none.

Keywords: cosine similarity, document embeddings, academia, economics, gender differences, labour force composition.

JEL classification codes: J16, J71, J82

1 Introduction

Academia is characterised by a global gender imbalance. Economics is one of the fields with lower female representation and socioeconomic diversity (Stansbury and Schultz, 2023). In Europe, the share of women working in economics in academic departments is overall 32%, and it becomes 27% in senior positions (Auriol *et al.*, 2022). Besides fairness concerns, the evidence that a more diverse workplace increases the level of productivity (for instance, by improving creativity performance) explains the need to understand why this gap persists and, eventually, which policies can address it. According to Bayer and Rouse (2016), the under-representation of women limits the questions asked and the identification of innovative perspectives through which familiar problems can be addressed.

This paper investigates the extent to which research similarity between senior and junior researchers is related to promotion in academia in economics and studies implications for gender diversity among faculty. The key idea we explore is that of self-image bias put forward in the psychology literature: people tend to assign greater weight to traits representing their strong points as compared to those representing their shortcomings (Hill *et al.*, 1988). Recently, Siniscalchi and Veronesi (2021) have developed a theoretical model where they incorporate self-image bias, suggesting that scholars promote scholars with more similar characteristics to their own, to explain women's under-representation in academia. Self-image bias, combined with heterogeneity by gender in field of study/ field of research (Chari and Goldsmith-Pinkham, 2017; Lundberg and Stearns, 2019; Belot *et al.*, 2023; Beneito *et al.*, 2021; Sierminska and Oaxaca, 2021) and with senior academics being mainly men, may be associated with the gender imbalance observed.

To address our research question, we need to observe the full set of candidates for positions and of seniors evaluating their applications. This information is clearly not easy to observe. The selection system for assistant professor positions in Italy helps to circumvent this data limitation issue, since selection is based on public calls advertising by local universities, which are required to publish a report on the selection process and the outcomes of the selection on their website. To measure research similarity we propose a novel measure based on text analysis, which can capture similarity in research characteristics such as research topics, methodologies used, policy relevance of the questions addressed, rather than mere research fields. We construct a dataset that covers the universe of job applications to public calls for tenure track assistant professorships in economics in Italy in the period 2014-2021, and we collect the abstracts of the papers of all candidates, selection committee members and faculty members of the departments issuing the calls. Specifically, using a Natural Language Processing tool (i.e., document embeddings), we compute the abstract similarity for each possible combination of publications of the candidate and the selection committee members and aggregate these similarity scores at the candidate-committee level, to examine how they relate to the selection outcome. We show that research similarity is positively associated with the probability of winning the selection and becoming a tenure track assistant professor, even when controlling for a rich set of candidate characteristics and candidate or call fixed effects. We also show that, although there is no gender gap in the average similarity between candidates and selection committee members, women and men differ in maximum similarity. Men are more likely than women to be very similar to one of the committee members, and this gender difference disappears when we focus only on female committee members. Last, we show that gender differences in (maximum) research similarity help explaining the gender gap in the probability of winning that we document when the probability of attending the interview in the last stage of the selection process is taken into account. We discuss the consistency of our results with similarity bias in male-dominated contexts contributing to the persistence of female under-representation in academia. We also highlight the narrowing of heterogeneity in research characteristics, with potential losses for the profession as a whole.

Our paper contributes to different strands of the literature. First, we add to the literature examining gender differences in research characteristics. Existing evidence shows that women do research in different fields of economics than men. Women are scarce in macro, finance and mathematical and quantitative methods, and more abundant in labour and other applied microeconomics fields (Chari and Goldsmith-Pinkham, 2017; Beneito *et al.*, 2021; Sierminska and Oaxaca, 2021). Greater disparities are found among academic economists than among graduate students (Sierminska and Oaxaca, 2021) and there is no evidence of significant changes over time (Lundberg and Stearns, 2019). We complement this literature by proposing a more granular measure of research characteristics besides fields of study, based on the application of NLP to paper abstracts. We show that research similarity is associated with the probability of winning, even when controlling for a rich set of candidate characteristics and fixed effects. We also show that the gender gap in the probability of winning is smaller when we control for research similarity. Second, we contribute to the literature on gender bias and under-representation of women in academia. Several papers document the existence of gender bias in academia, for instance, in teaching evaluations (Paredes *et al.*, 2023), in the publication process (Hengel, 2022; Sarsons, 2017), in citation patterns (Koffi, 2021), in reference letters (Baltrunaite *et al.*, 2023; Eberhardt *et al.*, 2023), and seminar behaviour (Dupas *et al.*, 2021). We test the presence of a specific type of bias, i.e. self-image or similarity bias, show its relationship with the outcome of the selection process and document gender differences in similarity. Our paper also complements the evidence on the role played by the gender of the evaluator in national assessments for promotion to associate and full professor, both in the Italian and in the Spanish context (De Paola and Scoppa, 2015; Bagues *et al.*, 2017). We show that the gender gap in similarity, which is positively related to the probability of winning, is driven by male members of the selection committee. Finally, our paper speaks to the literature using NLP to detect gender stereotypes and in-group bias (Ash *et al.*, 2023; Chen *et al.*, 2021; Koffi, forthcoming). More specifically, it is close to papers using NLP and word embeddings to measure gender bias (Ash *et al.*, 2024) and its influence on labour market performance (Baltrunaite *et al.*, 2023). We here adopt document embeddings as state-of-the-art framework in NLP to represent text as vectors and capture high levels of semantic complexity. The position of vectors in a multi-dimensional space can reveal closeness across publications under very many respects. We study similarity across abstracts to detect the presence of self-image bias and explore its relationship with the outcomes of selection processes, while at the same time supplying an enhanced measure of similarity/diversity in knowledge production, which can be used in other contexts, or to address different questions.

The paper is organised as follows. Section 2 and 3 describe the institutional setting and the dataset, respectively. Section 4 presents the methodology and validates our similarity measure. Section 5 provides descriptive evidence and discusses selection issues. Section 6 presents and discusses the results. Finally, Section 7 concludes.

2 Institutional framework

In Italy, the selection process for assistant professorships starts from a publicly advertised call. A department seeking to cover a (tenure track) assistant professor position decides the broad field of research of the call, indicates a full professor of the department who will be part of the selection committee (the internal member), together with two external members, who are randomly chosen from a restricted pool of professors from other universities that are indicated by the hiring department.[1] The selection process consists of multiple stages. In the first stage, the selection committee, whose composition is not public at the time candidates apply to the position, carries out the first screening and ranks candidates according to their CVs and publications, following pre-set criteria. These are decided upon by the committee, before the list of candidates applying for the position is known to them, and following broad rules decided at the University level (e.g., x points to be assigned to CVs and y points to be assigned to publications), in accordance with guidelines offered by the Ministry of University and Research. The selection committee writes a short evaluation report for all candidates, gives an overall assessment (e.g., excellent, very good, good, fair, below average), and drafts a shortlist with at least six candidates, who are invited to an interview with the selection committee. After the interview, the selection committee publishes a ranking of the candidates and indicates the winner. In some cases, information on the overall score assigned to each candidate and how it is split between CVs, publications and interview, is also published.

3 Data

By combining web-scraping techniques and manual retrieval, we build a novel dataset containing information on all candidates, members of selection committees and faculty of the hiring department for each call opened in Italy in the period 2014-2021 in the broad area of Economics, which is divided into Economics, Economic Policy, Public Economics, Econometrics, Applied Economics following the ministerial classification. Our dataset covers 237 calls for tenure track positions, involving 711 committee members and 2365 candidates.[2] Starting from the candidate dataset, it includes information on gender, publication records, university of the PhD, PhD

[1]Note that the profile of the ideal candidate can be defined only according to broad fields of research specified by Ministerial guidelines. These broad fields are Economics, Economic Policy, Public Economics, Econometrics, and Applied Economics. A finer definition of the field of research of the ideal candidate is not allowed (Law 30 December 2010, n. 240, `https://www.parlamento.it/parlam/leggi/10240l.htm`). Note also that some universities may not select members of the selection committee randomly. Since we do not exploit the random composition of the committee in our empirical strategy, this feature is not key in our setting. Also, it might be possible that a committee consists of external members only. This occurs, for instance, when there are no professors in the department opening the call that are hired in the broad field of the call.

[2]During the 2014-2021 period, 248 calls were issued. However, for 11 of them, we could not collect all the necessary information. Note that a candidate may participate in multiple calls. Our dataset includes 557 unique candidates.

graduation year, current occupation, the identification of the winning candidate. We also collect the publications of the candidates, and their abstracts in particular. In total, we have information on approximately 9500 publications of candidates. We then retrieve information on publication records and gender of each member of both the selection committees and the departments opening the call. In the collection of publications, we only consider faculty members who are economists and are assigned to the ministerial economic areas, which we listed above. In total, the dataset of members of selection committees and departments includes 1384 professors and approximately 33000 publications.

Our data come from three main sources. First, we use data from CINECA, which collects historical information on members of the departments. Second, the institutional websites of each Italian university have information on calls and their results, which allows us to construct the candidate side of the dataset. Finally, the publication data come from the Elsevier abstract and citation database SCOPUS.com (Elsevier, 2023), which provides information on author profiles, including affiliations, number of publications and their bibliographic data, references, and, importantly, the abstracts of the publications.[3] Using web-scraping techniques, for approximately 20% of the publications we are able to collect the corresponding JEL codes, which are not included in the data downloaded from SCOPUS.

Although our main analysis focuses only on senior (tenure track) assistant professorships, in order to shed light on the entire selection process we also collect the same type of data described above for calls for junior (non-tenure track) assistant professorships, as we will discuss in Section 5. Before applying for senior assistant professor positions, many candidates apply for junior assistant professor positions as a first step in the academic pipeline.

4 Methodology

We first describe the corpus construction, the methodology for text analysis and for the calculation of similarity, to then introduce the estimation equations.

[3]Scopus is an abstract and citation database offering extensive coverage of peer-reviewed literature and quality web sources in the fields of science, technology, medicine (STM), social sciences, and arts & humanities (A&H). It was launched in November 2004 and, currently, it has over 90.6 million core records and adds about 3 million new records annually (approximately 5,500 per day). The content in Scopus dates back to 1788, with references starting from the 1970s. Our data download took place in 2023 (Source: link).

4.1 Corpus construction

For each scholar in our dataset, i.e., candidates, members of selection committees and members of departments, we collect the abstract of all their publications. The overall number of publications and abstracts is 42506. We then consider all publications preceding the year of the call and, using text analysis, we calculate a measure of research similarity between candidate and members of the selection committee, and between candidate and members of the department opening the call.[4] The measures of research similarity are constructed by applying Natural Language Processing (NLP) techniques. As first step, we pre-process the texts of the abstracts of the papers by removing specific words related to copyright and editorial information, such as "Elsevier Ltd.", "Copyright", and "All rights reserved". Next, we represent each research paper using a document embedding of its abstract. A document embedding is a vector-based representation of a document, in this case, the abstract. The purpose of this representation is to capture the semantic meaning of the texts. Specifically, documents that share similar semantic characteristics will be represented by vectors that are closer to each other in a multidimensional space.

4.2 Document embeddings

To create document embeddings, we employ a specific technique called SentenceTransformers[5] (Reimers and Gurevych, 2019), which is a state-of-the-art framework for generating high-quality vector representations of sentences and documents. SentenceTransformers uses advanced deep learning models to encode the contextual information of the text, enabling the creation of meaningful and semantically rich document embeddings. It maps sentences and paragraphs to a 768 dimensional dense vector space. By leveraging the power of SentenceTransformers, our research similarity measures benefit from the latest advancements in NLP and provide accurate representations for comparing research papers based on their abstracts.[6]

Note that document embedding has consistently proven to be the most accurate model for semantic text similarity across various benchmark datasets (Conneau and Kiela, 2018). Semantic text similarity involves assessing the extent of *meaning* overlap between two texts. The performance of the model is assessed by how closely

[4] We exclude approximately 20 publications that are not in English and 4121 without an abstract.

[5] We employ `https://huggingface.co/sentence-transformers/all-mpnet-base-v2`

[6] A similar procedure to calculate topic similarity across publications is used in Koffi (forthcoming).

its output matches human judgement across a range of data sources, including news articles, Wikipedia, and user-generated content, with human annotators rating the similarity on a scale from 0 (completely dissimilar) to 5 (entirely equivalent). The close matching indicates that the similarity measure mainly captures content and topic similarity (and, by extension, methodology description) rather than variations in writing style or tone. However, despite the robust performance of SentenceTransformers representations across numerous NLP tasks (Conneau and Kiela, 2018), there is a limit in their interpretability, as they do not provide clear insights into the specific linguistic information they capture or the exact meanings they represent (Huber *et al.*, 2020).

Finally, this model provides distinct advantages over other approaches like TF-IDF.[7] Unlike TF-IDF, which relies on term frequency and inverse document frequency to represent documents, the SentenceTransformers framework utilises advanced machine learning techniques to generate the so-called contextual embedding. The latter allows for a more nuanced understanding of semantic similarity, capturing the meaning of words and phrases within their broader context. Consequently, SentenceTransformers excels in recognising semantic relationships, handling synonyms, and addressing polysemous words more effectively than TF-IDF. Notably, our methodology aligns with Koffi (forthcoming), who shows that SentenceTransformers matches most closely with human judgements in semantic similarity tasks, highlighting its superior performance in this area also when it comes to academic texts in the sciences and social sciences.

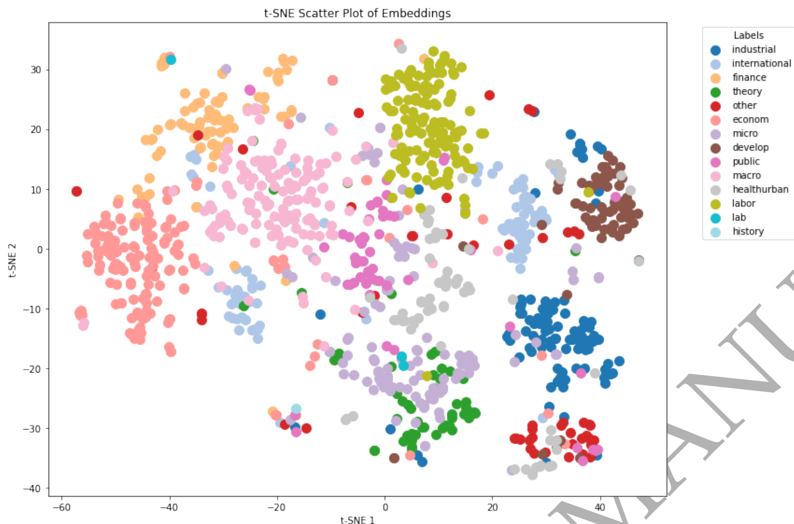
To understand the capability of document embeddings to represent abstracts, we visualise them in a two-dimensional space in Figure 1. This operation is possible by using the t-distributed Stochastic Neighbour Embedding (t-SNE) algorithm, which is an algorithm designed ad-hoc to reduce the dimensionality of document embeddings. The space proximity of two points in Figure 1 can be assimilable to the abstract similarity. Then, we colour each point with the respective research field. We use a sample of 1100 abstracts. The sample includes abstracts of papers that, according to their JEL codes, belong to only one of the fourteen research fields defined in Card and DellaVigna (2013).[8] This figure provides two main insights. First, although the embeddings were not trained to categorise research fields, they still reveal distinct field-related patterns. Second, as a continuous index, the embed-

[7]Biasi and Ma (2022) employ this methodology to analyse the education-innovation gap, looking at syllabi of University courses and knowledge, as reflected in research papers.

[8]See Section 4.4 for a description of these fields and of how JEL codes map into these fields.

dings capture richer information than field indicators and show that, even within field, there is heterogeneity in abstract representations. For instance, abstracts in macroeconomics that resemble those in public economics appear closer together in the graph. Conversely, abstracts in finance with less similarity to macroeconomics would be plotted farther away. Such spatial relationships offer a nuanced measure of content, mirroring the way we perceive concepts within a continuous spectrum.

Figure 1: t-SNE visualisation of the abstract embeddings by field of research



Notes. The Figure plots the t-distributed Stochastic Neighbor Embedding (t-SNE) visualisation of abstract embeddings and its relationship with fields of research. The Figure was generated using a sample of 1100 abstracts that, according to their JEL codes, can be univocally assigned to one of the fourteen research fields in the classification of Card and DellaVigna (2013). The fields are: experimental, history, development, urban/health, public, finance, international, IO, econometrics, labor, macro, micro, theory, other. Different colours represent different fields of research.

4.3 Indices of similarity

Once we have the vector representation of abstracts, we want to assess how similar they are. The similarity between two vectors (or embeddings) is traditionally determined using the cosine similarity. The cosine similarity index ranges between -1 and 1, where a smaller angle between two vectors (i.e., higher value of the index) indicates a higher degree of textual similarity. In our study, we use this metric to assess the similarity between the publication abstracts of each relevant combination of candidates and members of the selection committees (or departments). To summarise the results at the candidate-committee level, we aggregate the similarity scores obtained at the publication/abstract level. Specifically, we calculate the mean (Mean Sim) and the maximum (Max Sim) similarity between the publications of a candidate and those of the selection committee (or department) members.[9]

Figure A.1 illustrates the distribution of mean and maximum similarity between candidates and selection committee members. Given the texts we are considering are abstracts of economic papers, the mean and the maximum similarity are generally positive. The increase in the density of maximum similarity at 1 captures instances of co-authorship.

Figure A.2 shows two examples of pairs of abstracts with different cosine similarities. The first one is an example of high similarity (the cosine similarity between abstracts 1 and 2 is 0.93), while the second one is an example of low similarity (the cosine similarity between abstracts 3 and 4 is 0.008).

4.4 Validation of the measure

We perform several tests at the publication level, as well as at the scholar level to validate our measure of similarity based on document embeddings. In particular, we analyse the relationship between our similarity scores and JEL codes, journals where the papers were published, and primary field of research of the scholars in the sample.

First, at the publication level, we evaluate the correlation between our similarity score on the one side, and JEL codes and journals where the papers were published, on the other. In Figure A.3, we show that our similarity scores calculated across papers sharing at least one three-digit JEL code is 16 percentage points higher than across papers without common JEL codes. In addition, in Figure A.4, we show

[9]For each call, denote by C_{iz} the vector resulting from document embedding and representing publication z of candidate i, with $i = 1, \ldots, n_i$ and $z = 1, \ldots, n_z$. Similarly, denote by M_{kj} the vector resulting from document embedding of the publication j of committee member k, with $j = 1, \ldots, n_j$ and $k = 1, 2, 3$. We indicate by S_{zj}^{ik} the cosine similarity between publication z and publication j for each candidate i and each committee member k. The maximum similarity $MaxSim_i$ for candidate i corresponds to the highest cosine similarity calculated considering all z and j publications, and all committee members k in a given call, i.e., $MaxSim_i = \max S_{zj}^{ik}$ for all z, j, k. The mean similarity $MeanSim_i$ for candidate i is instead the average of the cosine similarities between publication z and publication j for all committee members k in a given call, i.e., $MeanSim_i = \frac{1}{3} \sum_k \frac{1}{n_z} \frac{1}{n_j} (\sum_z \sum_j S_{zj}^k)$.

how similarity changes with the number of JEL codes in common. It shows that one additional JEL code in common increases the average similarity score by 0.1-0.2 percentage points. Moreover, according to Figure A.5, papers published in the same journal are 15 percentage points more similar than papers published in different journals.[10]

Second, at the scholar level, we examine how the similarity between candidates and committee members varies based on the primary field of research of each scholar. To construct a variable for the primary field of research of a scholar, we first determine the field of each publication in our sample and then aggregate the information at the scholar level. Specifically, to determine the field of each publication in our dataset, we follow the field classification system used in Card and DellaVigna (2013), which assigns papers to fourteen not mutually exclusive fields, using the JEL codes of each paper.[11]

Since we have the information on the JEL codes only for approximately 20% of the publications in our sample, we use a topic model to predict the field of the publications for which we do not have a JEL code.[12] Specifically, we use a logistic regression as classification algorithm. Finally, we attribute as primary field to each scholar the one in which he/she has the majority of publications.[13] Figure 2 is a heat map, where each cell represents a combination of the candidate's and committee members' primary fields. The figure indicates that the mean and the max similarity are higher when the candidate and committee members share the same primary research field (i.e., in the diagonal of the matrix).[14]

Third, we analyse the relationship between similarity among candidates (rather than between candidates and committee members) and the broad field of the call. In Figure A.6, we show that the mean and maximum similarity among candidates applying to calls in the same broad field are 0.5 percentage points and 0.1 percentage

[10]Note that the analysis at the publication level is conducted by focusing on pairs of publication abstracts between candidates and members of the selection committee for calls which the candidate has participated to.

[11]The fields are: experimental, history, development, urban/health, public, finance, international, IO, econometrics, labor, macro, micro, theory, other. Note that a paper can be assigned to more than one field. See Card and DellaVigna (2013) for more details on the classification system.

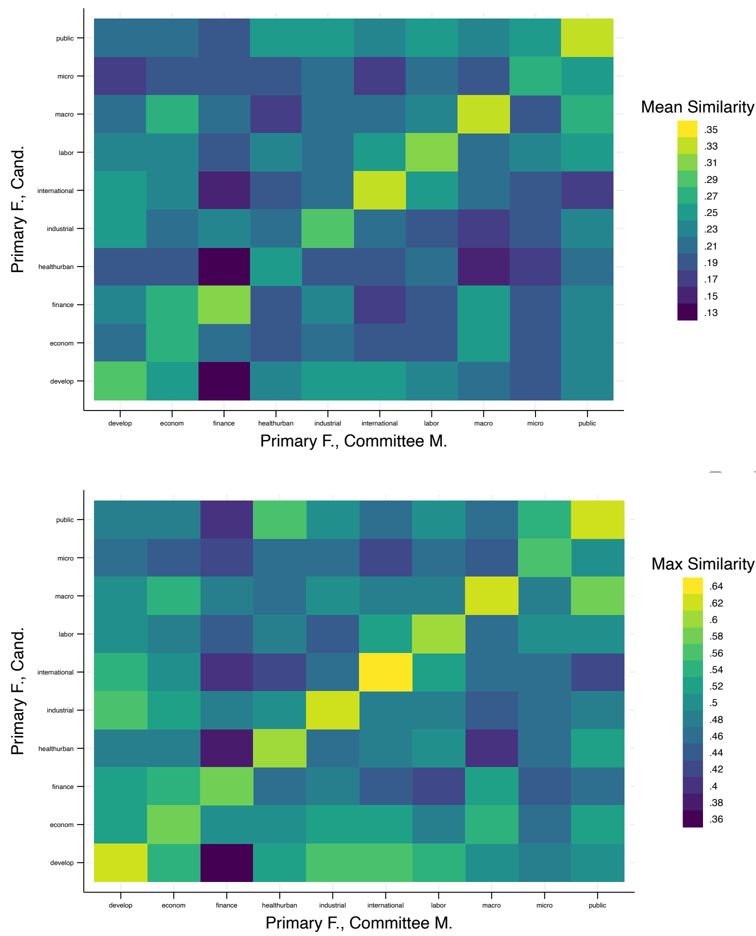
[12]This is because some journals do not include JEL codes in their papers or because they do not report them on the paper's preview, on which we apply our web scraping.

[13]If the main field of the scholar (i.e., the field with the largest number of publications) is "other", we consider the second field in terms of number of publications as the main field of the scholar.

[14]Note that no scholar has history as their main field, and only a few have theory or lab, which prevents these fields from being included in the heat map.

points higher, respectively, than those among candidates applying to calls in different broad fields.

Figure 2: Similarity distributions (Mean and Max) by primary field of research



Notes. The figure shows the distribution of the mean similarity and maximum similarity between candidates and members of selection committees by primary field of research. Each cell represents a combination of the candidate's and committee members' primary fields of research.

4.5 Estimation equations

Our main analysis consists of two steps. First, we test whether the similarity between candidates and selection committee members is associated with the probability of winning of the candidate. We then examine how similarity varies with the gender of the candidate.

Thus, we first estimate the following linear probability model:

$$Winner_{ijst} = \varphi_1 * DSimIndex_{ijst} + \varphi_2 X_i + Year_t + s_s + \varepsilon_{ijst} \tag{1}$$

where $Winner_{ijst}$ is a dummy equal to 1 if candidate i wins the selection of call j in the broad field s in year t. $DSimIndex_{ijst}$ is a dummy equal to 1 if the similarity index (Mean Sim or Max Sim) between candidate i and selection committee members for call j in year t and broad field s is above the 50th percentile.[15] X_i is a vector of candidate characteristics, namely, gender, years from PhD, whether the PhD was taken abroad, whether the candidate is employed abroad at the time of the call, number and quality of publications, whether the candidate is already working in the department launching the call *(Internal)* and whether he/she has co-authored at least a publication with a member of the selection committee. $Year_t$ and s_s are year and broad-field fixed effects, respectively. Finally, ε_{ijst} is the error term. In further specifications, we replace year and broad-field fixed effects with call or candidate fixed effects.

To analyse whether female and male candidates differ in terms of their similarity to the members of the selection committee, we estimate the following linear probability model:

$$DSimIndex_{ijst} = \varphi_1 Female_i + \varphi_2 X_i + Year_t + s_s + \varepsilon_{ijst} \tag{2}$$

where $DSimIndex_{ijst}$ is defined above, $Female_i$ is a dummy for female candidates, which captures gender differences in mean or maximum similarity, and all the other variables are defined as before (with the exception of gender that enters separately and it is not included in X_i). Likewise, in further specifications of the model, we replace year and broad-field fixed effects with call fixed effects.

[15]Figure A.7 reports the average winning probability for each quintile of the similarity distribution, for both the mean and the max similarity indices. Candidates in the highest quintiles have a significantly higher probability of winning. Thus, the Figure shows the non linearity of the relationship between our similarity indices and the probability of winning, which justifies the choice of using a dummy instead of a continuous variable. We assess the robustness of our results to the adoption of a continuous measure of similarity in Section 6.1.1.

5 Descriptive evidence

In this section, we present summary statistics and discuss selection by gender into applying to senior assistant professor positions and participating in job interviews, to examine the existence of a gender gap in the probability of winning and lay the ground to discuss the role of research similarity.

5.1 Descriptive statistics

Table 1 shows that our dataset includes 2365 observations. The share of women is 36%. The probability for a candidate of winning a selection is 10%. On average, the share of women in committees is 32%.[16] Each call has, on average, 16 candidates.

Table 1: Summary statistics

Calls for senior assistant professorships

Variable	*Mean*	*Sd*	*Obs.*
Female	0.356	0.479	2365
Winner	0.101	0.301	2365
PhD Abroad	0.240	0.427	2365
Currently Abroad	0.274	0.446	2365
Years from PhD	7.175	3.092	2365
N cand/call	15.821	8.794	2365
Share women in the Committee	0.316	0.229	2365

Notes. The table provides summary statistics for the following variables: share of females, probability of winning, share of candidates with a PhD abroad or currently abroad, average number of years from PhD, average number of candidates per call and average share of women in selection committees for senior assistant professorships. Years: 2014-2021.

In Table 2, we report summary statistics by gender of the candidate, focusing on observable characteristics, including the publication record, and on similarity with selection committee members. The results of a t-test show that, although the

[16] In Figure A.8, we show the dynamics of the share of women in selection committees, which displays limited variation over time.

Table 2: Summary statistics: Differences by gender

Candidates for senior assistant professorships

Panel 1: Characteristics					
Variable	*Men*	*Women*	*T-STAT*	*Diff*	*p-value*
Winner	0.10	0.10	-0.18	0.00	0.86
Shortlisted	0.51	0.50	0.58	0.01	0.56
Present	0.57	0.65	-2.56	-0.08	**0.01**
PhD Abroad	0.23	0.26	-1.41	-0.03	0.16
Currently Abroad	0.28	0.25	1.59	0.03	0.11
Years from PhD	7.08	7.34	-1.94	-0.26	**0.05**
Internal Candidate	0.04	0.10	-5.10	-0.06	**0.00**
Co-author	0.02	0.04	-1.65	-0.01	0.10
Panel 2: Publication Record					
Variable	*Men*	*Women*	*T-STAT*	*Diff*	*p-value*
At least one Top 6	0.01	0.02	-1.47	-0.01	0.14
N pubs in A+	0.15	0.25	-4.54	-0.11	**0.00**
N pubs in A	6.30	5.84	2.76	0.46	**0.01**
N pubs	9.88	8.96	3.09	0.92	**0.00**
At least one interdisciplinary	0.03	0.02	1.48	0.01	0.14
Panel 3: Similarity					
Variable	*Men*	*Women*	*T-STAT*	*Diff*	*p-value*
Mean Sim with Committee	0.218	0.222	-1.386	-0.004	0.166
Max Sim with Committee	0.594	0.590	0.635	0.004	0.525

Notes. The table reports summary statistics and t-tests by gender of the candidates for the following variables: probability of winning the selection, probability of being shortlisted, probability of being present at the interview, share of those with a PhD abroad, share of those working abroad at the time of the selection, average number of years from PhD, share of internal candidates, share of candidates with a co-author in the selection committee, share of those with at least one Top 6 publication, average number of A+ publications, average number of A publications, share of those with at least one interdisciplinary publication, mean and max similarity with the committee. Top 6 journals are AER, QJE, JPE, REStud, Econometrica, JF; A+ journals are AEJs, EJ, JEEA, Rand, JPubE, JME, RFS, JEc, JOLE, JHE, QE, JTE, JDE, JIE, Theor. Econ.; A journals are defined according to the ANVUR classification.

probability of winning the selection or being shortlisted for the interview does not differ by gender, women seem to be more qualified candidates: they have a higher number of publications in A+ journals, and they are also more senior, since more years passed from the PhD defence to the time of the call. On the other hand, men have, on average, a higher number of A publications and total publications.[17] Interestingly, women are more likely than men to be present at the interview, when shortlisted, and more likely to be internal candidates. Finally, in Panel 3 there is no evidence of significant gender differences in terms of similarity with the selection committee.

We further explore differences in similarity in Figures A.9-A.10, where we show the distributions of our similarity indices by gender of the candidate and gender of the committee members. Interestingly, while there are no clear differences by gender in the similarity distributions when we focus on female committee members, the distribution of the maximum similarity with male committee members for female candidates appears to be left-shifted, compared to that for male candidates, suggesting that female candidates have lower values of maximum similarity with male committee members, compared to male candidates.[18] In Figure A.11, we provide the distributions for similarity among candidates (instead of those between candidates and members of the committees), separately for female and male candidates. The distributions appear quite similar, which is not in line with the hypothesis that female candidates do research on a smaller group of topics compared to male ones. In Figure A.12, we distinguish candidates applying to the same call and candidates applying to other calls in the same broad field and plot the distributions of mean and maximum similarity: the two distributions are not statistically different according to the Kolmogorov-Smirnov test, except for economics, economic policy, and public economics, where the distributions for the mean similarity are significantly different. This evidence supports the hypothesis that candidates do not select into calls based on the similarity with the committee, which is consistent with them not observing the composition of the committee at the time of application, as we will further discuss in Section 6.4.

[17]Top 6 journals are AER, QJE, JPE, REStud, Econometrica, JF; A+ journals are AEJs, EJ, JEEA, Rand, JPubE, JME, RFS, JEc, JOLE, JHR, QE, JTE, JDE, JIE, Theor. Econ.; A journals are defined according to the ANVUR - National Agency for the Evaluation of University and Research - classification (link).

[18]The distributions for the maximum similarity with the male members of the committee are different by gender according to the Kolmogorov- Smirnov test.

5.2 The selection into the pool of candidates

According to the summary statistics reported in Tables 1 and 2, women are under-represented among candidates for senior assistant professorships. Moreover, female candidates are characterised by a higher academic quality (i.e., they have a higher number of highly ranked publications) and a higher academic age (i.e., they have a higher number of years from the PhD graduation at the time of the application) than their male counterparts. This suggests that the selection of researchers in the pool of candidates for tenure track assistant professorships may operate differently for women and men.

To investigate such differences, we examine the role of self-selection in applying for tenure track assistant professorships and gender differences in the probability of winning at an earlier stage of the academic career ladder.

We proceed as follows. First, we explore whether female and male researchers differ in terms of their probability of applying for a senior assistant professorship. To this end, we construct a pseudo dataset at the candidate/call level in which, for each candidate applying for at least one call in a given year, we add observations also for the other calls in the same or subsequent years and in the same broad field, for which the candidate has not applied, and generate a dummy equal to 1 if the candidate has applied for that specific call and 0 otherwise.[19] Using this dummy as dependent variable, we estimate an equation akin to Equation 1 (without including the research similarity dummy) and investigate the role of gender in explaining our outcome of interest, i.e., the application probability. The results in Table A.1 show that, conditional on observable characteristics and on the inclusion of fixed effects, the application probability is around 0.9 percentage points lower for female candidates compared to male candidates, which corresponds to 12% of the average application probability $\overline{Y}$ in the sample.[20]

Second, we investigate whether, in addition to self-selection, the lower proportion of women among candidates for senior assistant professorships depends on the lower probability of success of female candidates at the previous stage of the tenure process, i.e., the selection for junior assistant professorships. To do so, we collect information

[19]For each candidate, we add observations until the year in which the candidate wins a selection or, if the candidate never wins a selection, until 2021, the last year of our period of analysis.

[20]We check whether this is driven by the fact that women are more likely to leave academia after a previous negative outcome in the competition. Specifically, we generate a dummy equal to 1 if, in the previous year, the candidate did not win any call he/she had applied for, and analyse whether the correlation between this dummy and the probability of applying differs by gender of the candidate. We do not find evidence supporting this hypothesis.

on the universe of calls and candidates for junior assistant professorships. Descriptive statistics on this dataset are provided in Tables A.2 and A.3. During the 2015-2021 period, we have identified 162 calls and 954 candidates for which we have gathered all the necessary information.[21] As for candidates for tenure track assistant professorships, we observe whether the candidate is the winner of the competition, the year of the PhD and whether she/he has received the PhD abroad, whether she/he is abroad at the time of the application, and the publication record. Interestingly, Table A.2 shows that women represent a higher proportion compared to the pool of applicants for senior positions (40% vs 36%). Moreover, female candidates do not appear significantly different in terms of observable characteristics compared to male candidates: in particular, there is no evidence of a statistically significant difference in the number of publications in A+ journal, while the average number of publications in A journal is slightly higher for men (Table A.3). Yet, female candidates are less likely to win.

We therefore examine empirically whether female candidates have a lower probability of winning the selection than male candidates. To do this, we construct a dummy equal to 1 if the candidate wins the competition, and equal to 0 otherwise.[22] According to Table A.4, the probability of winning the selection is much lower (5.7-7.4 percentage points or 33-42%) for female than for male candidates, even conditional on their publication record.

These results suggest that the evidence that women make up a lower percentage of candidates for senior assistant professorships and that those participating in the selection are better candidates than their male counterparts is consistent with two pieces of evidence. First, women are less likely to apply than men, maybe because they prefer to gain more experience and publications before applying for a senior position,[23] or because they are less mobile (Le Barbanchon *et al.*, 2021). Second, there is a gender gap in the probability of winning at the entry-level of the profession, which reduces the participation of women in calls for senior assistant professorship and explains why those female researchers who do participate seem to come from the upper tail of the quality distribution.

[21]Note that a candidate may participate in multiple calls. Our dataset includes 463 unique candidates for junior assistant professorships. Note also that we have not identified any call in 2014.

[22]Note that the way selection procedures for junior assistant professorships work is the same as that for senior assistant professorships that we have described in Section 2.

[23]This is line with the literature on gender differences in competition, see for instance Niederle and Vesterlund (2007) and Gneezy and Rustichini (2004).

5.3 Gender gap in the probability of winning and selection bias

Before turning to our main empirical analysis, we examine whether there is a gender gap in the probability of success in the competition for senior assistant professorships, ignoring for the moment the role played by research similarity. To do so, we use a specification similar to Equation 1, without including the similarity dummy.

The results are reported in Table A.5. While our main dependent variable is the probability of winning the competition (column 1), we also look at the probability of being shortlisted (column 2) and the probability of being present at the interview, if shortlisted (column 3). According to the results reported in the table, the coefficient of the female dummy is negative in the first two columns, although not statistically significant, while it is positive and significant in column 3. This indicates that female candidates are more likely than male candidates to attend the interview if shortlisted (the participation probability is 6 percentage points, or 10% higher for women than for men), in line with the descriptive evidence from Table 2.

As the probability of being present at the interview, which also depends on the probability of being shortlisted, strongly influences the probability of winning the competition, we correct for this selection bias in our analysis of the gender gap in the probability of winning. Specifically, similar to the Heckman selection model, we implement a three-stage procedure. In the first stage, we estimate the probability of being shortlisted using a linear probability model where the dependent variable is a dummy equal to 1 if the candidate is selected for the interview and equal to 0 otherwise, and as covariates all the observable candidate characteristics listed in Section 4, plus a variable capturing the number of applications the candidate makes per year. Then, in the second stage, we regress the probability of being present at the interview on the same covariates, plus the predicted value of the probability of being shortlisted. Finally, in the last step, we estimate a linear probability model similar to Equation 1, extended to include our estimate of the probability of participating in the interview.

The results are shown in Table 3. Interestingly, the coefficient of the gender dummy is now statistically significant, negative, and bigger in magnitude compared to the one in Table A.5, column 1. This provides evidence that women are characterised by a lower probability of winning the competition for senior assistant professor positions than men. As expected, the coefficient of the probability of being present is positive and highly significant.

Table 3: Gender gap in the probability of winning, LPM

VARIABLES	(1) Winner	(2) Winner
Female	-0.077***	-0.071***
	(0.016)	(0.017)
PhD Abroad	0.008	0.007
	(0.016)	(0.017)
Abroad	-0.075***	-0.051***
	(0.016)	(0.017)
Years from PhD	-0.004*	-0.006***
	(0.002)	(0.002)
Internal Candidate	0.012	0.007
	(0.044)	(0.045)
Co-author	0.083	0.059
	(0.062)	(0.064)
At least one Top 6	-0.060	-0.091
	(0.066)	(0.074)
N pubs in A+	0.060***	0.071***
	(0.015)	(0.015)
At least one Interd.	-0.027	-0.021
	(0.043)	(0.048)
N pubs in A	0.028***	0.029***
	(0.003)	(0.004)
N pubs	-0.007***	-0.008***
	(0.001)	(0.001)
Pr (Present)	1.007***	1.068***
	(0.148)	(0.163)
$\overline{Y}$	0.10	0.10
Observations	2,365	2,365
R-squared	0.088	0.168
Call FE	No	Yes
Year FE	Yes	No
Broad Field FE	Yes	No

Notes. Dependent variable: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. Pr(present) is a OLS estimate of the probability of being present at the interview. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6 Results

In the previous section, we showed that there are important gender differences in selection for senior assistant professorships. Women are less likely than men to apply for senior assistant professorships. However, when they do apply, they are more likely to be present at the interview, if they are shortlisted. Although women appear to be better candidates than men on some dimensions, they are less likely to be successful than their male counterparts and win a senior assistant professorship.

In this section, we first consider whether research similarity is associated with the probability of winning and conduct a series of robustness tests and heterogeneity analyses to further explore the role of similarity in the probability of success. We then examine whether there is evidence of gender differences in research similarity that can explain the gender gap in the probability of winning we have documented in the previous section.

6.1 The effect of research similarity on the probability of winning

Table 4 reports the results of the estimation of Equation 1. While columns 1-3 use the mean similarity dummy as key explanatory variable, columns 4-6 investigate the role played by the maximum similarity dummy. Columns 1 and 4 show the results of the specification with broad-field and year fixed effects; columns 2 and 5 those with call fixed effects. Finally, columns 3 and 6 incorporate year and candidate fixed effects.

The results show that similarity is positively related to the probability of winning. Candidates for whom the dummy based on average similarity is equal to 1 are 6 percentage points more likely to win than those for whom the dummy is equal to 0. The coefficient is also positive and significant – though slightly smaller in magnitude – when we use more demanding specifications and include call or candidate and year fixed effects. The effect of the maximum similarity dummy is similar in magnitude and significance in all specifications. Consistent with the results in the previous section, when we do not control for the probability of attending the interview (equation 1), the female dummy is negative but insignificant, suggesting that women and men do not differ in their chances of becoming senior assistant professors. It is also worth noting that the effect of the similarity indices is even larger than that of an additional A+ publication (in columns 1-2, 4-5). Besides the positive effect of high quality publications on the probability of winning, it is interesting to note the positive and large effect of having a co-author on the selection committee and of being an internal candidate.

6.1.1 Robustness checks and heterogeneity analysis

We run several robustness checks. First, we check that our results are robust to controlling for finer fields of research of candidates and members of the selection committee, in particular the internal one, who may play a more relevant role in the selection process. Specifically, in our regression, we include dummies for the primary field of research of the internal member of the committee (Table 5, Panel 1), for the primary field of research of the candidate (Table 5, Panel 2), and for candidates sharing the same primary field of the internal member of the committee (Table 5, Panel 3). The coefficients of our similarity indices remain always positive and statistically significant. This indicates that, even after accounting for candidates working in the same field as the internal member of the committee, our similarity indices account for part of the variation in the candidates' likelihood of winning.[24]

[24]The smaller number of observations in Panels 1 and 3 is explained by the fact that 43 calls lack internal members on the committee.

Table 4: The role of research similarity in the probability of winning, LPM

VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Dummy Similarity	0.061***	0.058***	0.048***	0.056***	0.057***	0.046***
	(0.012)	(0.013)	(0.016)	(0.012)	(0.013)	(0.014)
Female	-0.014	-0.006		-0.012	-0.003	
	(0.013)	(0.013)		(0.013)	(0.013)	
PhD Abroad	0.021	0.020		0.022	0.020	
	(0.016)	(0.017)		(0.016)	(0.016)	
Abroad	-0.016	0.009	0.012	-0.011	0.016	0.015
	(0.013)	(0.015)	(0.019)	(0.013)	(0.015)	(0.019)
Years from PhD	0.001	-0.000	0.064*	0.001	0.000	0.065*
	(0.002)	(0.002)	(0.038)	(0.002)	(0.002)	(0.039)
Internal Candidate	0.180***	0.188***	0.097**	0.186***	0.192***	0.103**
	(0.037)	(0.037)	(0.041)	(0.037)	(0.037)	(0.041)
Co-author	0.223***	0.214***	0.157**	0.217***	0.204***	0.151**
	(0.057)	(0.058)	(0.062)	(0.058)	(0.059)	(0.062)
At least one Top 6	0.037	0.014	-0.070*	0.041	0.013	-0.069*
	(0.066)	(0.072)	(0.036)	(0.066)	(0.072)	(0.036)
N pubs in A+	0.033**	0.040***	0.128**	0.033**	0.041***	0.128**
	(0.015)	(0.015)	(0.062)	(0.015)	(0.015)	(0.062)
At least one Interd.	0.040	0.047	-0.128***	0.037	0.044	-0.089*
	(0.042)	(0.047)	(0.046)	(0.042)	(0.048)	(0.048)
N pubs in A	0.011***	0.012***	0.014	0.011***	0.011***	0.011
	(0.002)	(0.002)	(0.012)	(0.002)	(0.002)	(0.012)
N pubs	-0.003***	-0.003***	-0.006	-0.003***	-0.004***	-0.005
	(0.001)	(0.001)	(0.009)	(0.001)	(0.001)	(0.009)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.083	0.160	0.450	0.081	0.160	0.450
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variable: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: The role of similarity in the probability of winning probability, LPM

Controlling for the primary field of research of internal committee members and candidates

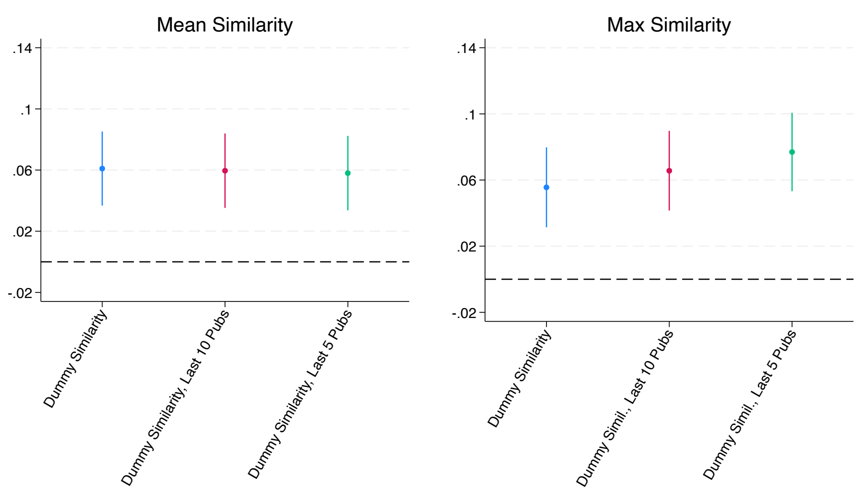
VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Panel 1: Controlling for the primary field of research of the internal member						
Dummy Similarity	0.059***		0.047***	0.051***		0.039***
	(0.013)		(0.016)	(0.013)		(0.014)
Observations	2,077		2,077	2,077		2,077
R-squared	0.090		0.471	0.088		0.470
Panel 2: Controlling for the primary field of research of the candidate						
Dummy Similarity	0.067***	0.064***		0.058***	0.059***	
	(0.013)	(0.014)		(0.012)	(0.013)	
Observations	2,365	2,365		2,365	2,365	
R-squared	0.088	0.165		0.086	0.164	
Panel 3: Including the same field of research dummy						
Dummy Similarity	0.055***	0.055***	0.042**	0.048***	0.049***	0.037***
	(0.013)	(0.013)	(0.017)	(0.013)	(0.014)	(0.015)
Same field Dummy	0.025	0.030	0.026	0.030	0.036*	0.029
	(0.021)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)
Observations	2,077	2,077	2,077	2,077	2,077	2,077
R-squared	0.089	0.151	0.469	0.087	0.150	0.469
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variable: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. In all panels our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, average number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, share of those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Second, we analyse whether the effect of research similarity on the probability of winning is robust to a change in the measure of research similarity. Specifically, we construct our two similarity indices focusing only on the most recent publications of the committee members. Specifically, the 10 and 5 most recent publications. We recall that a publication is included in the similarity measure only when it precedes the time of the call. Figure 3 shows the results and indicates that this change in the construction of our measures does not affect the results. If any difference exists, the effect of the maximum similarity increases when we focus on the last 5 publications.

Figure 3: The role of similarity in the probability of winning

Alternative measures of similarity



Notes. The figure shows the coefficients of the two similarity dummies in the estimation of Equation 1 (mean or max similarity) for alternative measures of similarity (considering all the publications of the selection committee members, only the last 10 most recent publications, and only the last 5 most recent publications). A publication is included if it was published before the date of the call.

As a third robustness check, we test whether our measures of similarity are simply capturing common networks. To investigate this issue, we rerun our regressions including among the controls a dummy equal to 1 if the candidate and one of the committee members have a common co-author, and equal to 0 otherwise. The results are reported in Table 6, Panel 1, and suggest that, although common networks play an important role in influencing the probability of winning, the relationship between research similarity and the probability of winning is robust to the inclusion of this additional control.[25]

In addition, we check the robustness of our results by including a measure of the impact/quality of the candidate's publications, namely, the average number of citations per paper. The results are provided in Table 6, Panel 2. The inclusion of this variable does not affect the results. We also examine the role of research similarity between candidates and members of departments opening the call and study whether our results are robust to including this measure of similarity in the regression. Table 6, Panel 3, shows that the coefficients of our similarity dummies are still positive and statistically significant, while the similarity indices computed with respect to department members are not statistically significant, with the exception of column 6.

The results show that being an internal candidate strongly influences the probability of winning the selection. In order to address the concern that this might indicate the presence of some private information sharing between the internal candidate and the department, we check that the effect of similarity still holds when we focus only on those calls where there are no internal candidates. The results are provided in Table 7, Panel 1. The effect of our similarity measures now appear to be even stronger. Our results also hold when we focus on the similarity between the candidate and only the external members of the committee (Table 7, Panel 2).

In a further check, we change our dependent variable and analyse the influence of research similarity between candidates and members of selection committees on the probability of being shortlisted for the interview, rather than on the winning probability.[26] The results in Table A.6, Panel 1 confirm the robustness of the effect of similarity: our similarity dummies are associated with an increase in the probability of being shortlisted for the interview, ranging from 14 to 7 percentage points. In Table A.6, Panel 2, we replicate the same analysis for the probability of being present at the interview and show that similarity does not play any role,[27] whereas female candidates are more likely to be present.

We also check that our results are robust to using a continuous variable for research similarity instead of the mean and maximum similarity dummies. The results are reported in Table A.7, Panel 1 and confirm the robustness of our results. Finally, we confirm that our results are robust to using a probit model instead of the

[25] In our dataset, the probability of having a co-author in common is 14%.

[26] This allows us to further address the concern that the probability of winning can be endogenous to the decision of those candidates that have been admitted to the interview to attend it.

[27] The coefficient of the dummy similarity is statistically significant only in column 1, but it is negative.

linear probability model.[28]

After having checked robustness, we conduct two heterogeneity analyses. First, we examine whether the role played by the similarity between candidates and the selection committee members varies by gender. To address this question, we add an interaction term between the female dummy and the mean/max similarity dummy to the specification in Equation 1. The results of this new specification are included in Table 8 and show that the association of similarity with the winning probability does not vary by gender. Similarity is positively related with the probability of winning for both female and male candidates.[29] Table A.7, Panel 2 shows the robustness of this result using the continuous similarity variable.

[28]The results are available upon request.

[29]Note that we do not examine how the effect varies with the proportion of female members on the selection committee, as there is not enough variation in the gender composition of the committee to explore this. As shown in Table 1, women on average represent 32% of the members of the committees (less than 1 in 3 members) and the standard deviation is quite low. Figure A.8 shows that the change over time is also limited.

Table 6: The role of similarity in the probability of winning probability, LPM

Robustness checks I

VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Panel 1: Controlling for common network						
Dummy Similarity	0.051***	0.048***	0.043**	0.044***	0.047***	0.042***
	(0.012)	(0.013)	(0.016)	(0.012)	(0.013)	(0.014)
Common Network	0.090***	0.090***	0.041*	0.091***	0.090***	0.041*
	(0.022)	(0.022)	(0.024)	(0.022)	(0.022)	(0.024)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.093	0.169	0.452	0.091	0.169	0.452
Panel 2: Controlling for citations						
Dummy Similarity	0.061***	0.058***	0.048***	0.056***	0.057***	0.046***
	(0.012)	(0.013)	(0.016)	(0.012)	(0.013)	(0.014)
Average N Citations	0.000	0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.002)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.083	0.160	0.450	0.081	0.160	0.450
Panel 3: Controlling for similarity with the department						
Dummy Similarity	0.060***	0.061***	0.045***	0.053***	0.051***	0.047***
	(0.014)	(0.015)	(0.016)	(0.013)	(0.013)	(0.014)
Dummy Similarity Depart.	-0.009	-0.016	-0.001	0.001	0.014	-0.030**
	(0.014)	(0.016)	(0.018)	(0.012)	(0.015)	(0.014)
Observations	2,341	2,341	2,341	2,341	2,341	2,341
R-squared	0.081	0.155	0.448	0.079	0.155	0.449
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variable: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. *Common network* is equal to 1 if the candidate and a member of the committee have a common co-author. *Citations* measures the average number of citation per publication of the candidate up to 2023. *DummySimDepart* measures the similarity between the candidate and the economics faculty of the department opening the call. In all panels our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, average number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, share of those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: The role of similarity in the probability of winning, LPM

Robustness checks II

VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Panel 1: Only calls with no internal candidates						
Dummy Similarity	0.064***	0.063***	0.075***	0.057***	0.065***	0.064***
	(0.018)	(0.020)	(0.025)	(0.018)	(0.019)	(0.022)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
Observations	1,313	1,313	1,313	1,313	1,313	1,313
R-squared	0.059	0.135	0.514	0.056	0.135	0.512
Panel 2: Similarity only with external committee members						
Dummy Similarity	0.052***	0.052***	0.040***	0.050***	0.049***	0.045***
	(0.013)	(0.013)	(0.015)	(0.013)	(0.014)	(0.014)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
Observations	2,299	2,299	2,299	2,299	2,299	2,299
R-squared	0.082	0.161	0.456	0.081	0.160	0.457
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variable: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. In all panels our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

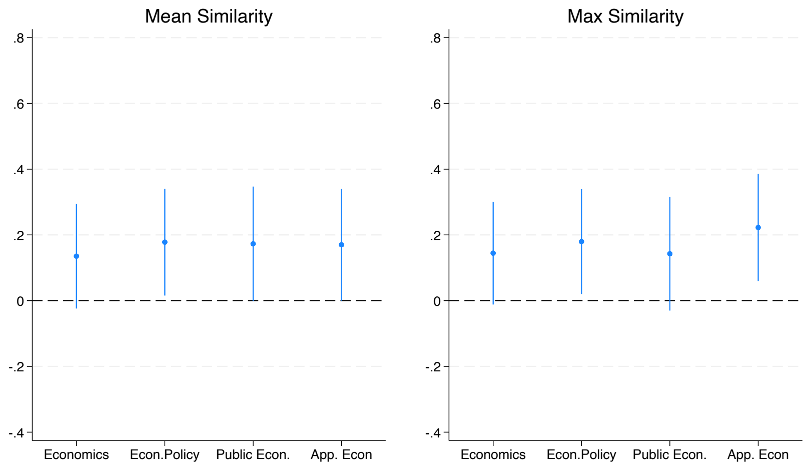
Table 8: The role of similarity in the probability of winning, LPM

Heterogeneity by candidate gender

VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Dummy Similarity	0.066***	0.060***	0.054***	0.050***	0.049***	0.044***
	(0.015)	(0.016)	(0.019)	(0.015)	(0.016)	(0.016)
Dummy Similarity*Female	-0.016	-0.007	-0.018	0.015	0.024	0.007
	(0.025)	(0.027)	(0.034)	(0.025)	(0.027)	(0.030)
Female	-0.006	-0.003		-0.019	-0.015	
	(0.016)	(0.017)		(0.015)	(0.015)	
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.083	0.160	0.450	0.081	0.160	0.450
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variables: Probability of winning a senior assistant professorship in economics. Estimates from a Linear Probability Model. In all panels our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 *** p<0.01, ** p<0.05, * p<0.1

Figure 4: The role of similarity in the probability of winning by broad field of the call



Notes. The graphs shows the coefficients of the interaction terms between the similarity dummies and the broad field dummies (Equation 1). Econometrics is the omitted category.

Last, we investigate whether the association between similarity and the probability of winning varies by broad fields (Economics, Economic Policy, Public Economics, Applied Economics, Econometrics). We include in Equation 1 interaction terms between the mean and max similarity dummies, respectively, and broad-field dummies. In Figure 4, we plot the coefficients of the interaction terms. Econometrics is the omitted category. The figure shows that there are small differences across macro-fields with respect to the role of the mean and max similarity.

6.2 Gender differences in research similarity

We now discuss the results of the estimation of Equation 2, to see whether female and male candidates differ in terms of mean and maximum similarity with the selection committee. The results are reported in Table 9. Columns 1-3 use as dependent variable the dummy for the mean similarity to the selection committee, while columns 4-6 the dummy for the maximum similarity to the selection committee. Columns 2 and 5 include year and broad field fixed effects, while columns 3 and 6 include call fixed effects.

Table 9: Gender differences in similarity, LPM

	Mean	Mean	Mean	Max	Max	Max
	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.002	0.002	0.034	-0.036*	-0.041**	-0.026
	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.022)
PhD Abroad	-0.018	-0.005	-0.008	-0.011	-0.010	-0.001
	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.025)
Abroad	0.027	0.035	0.068***	-0.074***	-0.067***	-0.053**
	(0.024)	(0.024)	(0.026)	(0.023)	(0.023)	(0.024)
Years from PhD	0.004	0.005	0.006*	0.000	0.000	0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Internal Candidate	0.071*	0.064	0.068	-0.017	-0.025	0.000
	(0.043)	(0.042)	(0.044)	(0.041)	(0.040)	(0.040)
Co-author	0.356***	0.333***	0.302***	0.487***	0.469***	0.484***
	(0.045)	(0.044)	(0.051)	(0.019)	(0.025)	(0.035)
At least one Top 6	-0.039	-0.019	-0.039	-0.134	-0.102	-0.023
	(0.095)	(0.093)	(0.085)	(0.088)	(0.085)	(0.086)
N pubs in A+	0.027	0.031	0.047**	0.025	0.031*	0.033
	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.020)
At least one Interd.	-0.061	-0.049	-0.031	-0.027	0.004	0.022
	(0.066)	(0.065)	(0.064)	(0.067)	(0.067)	(0.067)
N pubs in A	0.016***	0.019***	0.021***	0.023***	0.027***	0.027***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
N pubs	-0.014***	-0.015***	-0.017***	-0.000	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\overline{Y}$	0.5	0.5	0.5	0.5	0.5	0.5
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.036	0.071	0.222	0.073	0.116	0.232
Call FE	No	No	Yes	No	No	Yes
Year FE	No	Yes	No	No	Yes	No
Broad Field FE	No	Yes	No	No	Yes	No

Notes. Dependent variables: Mean/Maximum Similarity between the candidate and members of the committee. Estimates from a Linear Probability Model. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Gender differences in similarity, LPM,

Controlling for the primary field of research of the candidate

	Mean (1)	Mean (2)	Mean (3)	Max (4)	Max (5)	Max (6)
Female	-0.015	-0.016	0.010	-0.054**	-0.057**	-0.041*
	(0.023)	(0.022)	(0.023)	(0.023)	(0.022)	(0.023)
$\overline{Y}$	0.5	0.5	0.5	0.5	0.5	0.5
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.083	0.113	0.257	0.089	0.128	0.242
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	No	Yes	No	No	Yes
Year FE	No	Yes	No	No	Yes	No
Broad Field FE	No	Yes	No	No	Yes	No

Notes. Dependent variables: Mean/Maximum Similarity between the Candidate and Members of the committee. Estimates from a Linear Probability Model. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals, and dummies for the primary field of research of the candidate. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

The coefficient of the female dummy is not statistically different from 0 in the first three columns, but it becomes larger in size and statistically significant in columns 4-5. This suggests that, although female and male candidates do not differ in terms of mean similarity, male candidates are more likely than female candidates to be very similar to a selection committee member. The probability that the dummy variable based on the maximum similarity is equal to 1 is 3-4 percentage points or 7-8% larger for male candidates than for female candidates. When looking at the other controls, being abroad relates negatively to maximum similarity, indicating that candidates applying from abroad are less close in terms of research to selection committee members, in most cases based in Italian universities. Similar, if not stronger, results are found in Table 10, where we also control for the primary field of research of the candidate.

We now examine if the gender composition of the selection committee matters for the gender gap in maximum similarity we have shown. We estimate again equation 2, focusing first only on the female members, and then only on the male members of the committees. The results are provided in Table 11. Interestingly, we find that the gender gap in the maximum similarity disappears when we look only at female committee members (Table 11, Panel 1), while it is larger when we focus only on male members (Table 11, Panel 2). This supports the hypothesis that female candidates are less likely to be very similar to one of the committee members because selection committees are predominantly composed of men.[30]

Finally, we explore whether the gender gap in similarity varies across broad fields. As before, we add to Equation 2 an interaction term between the female dummy and the broad-field dummies. The results, reported in Figure 5, show that, while the gender gap in maximum similarity does not vary by macro-field, the gender gap in mean similarity is larger for economics and public economics, compared to econometrics, which is the omitted category, possibly indicating more heterogeneity in research interests and methodologies in the broad fields of economics and public economics.

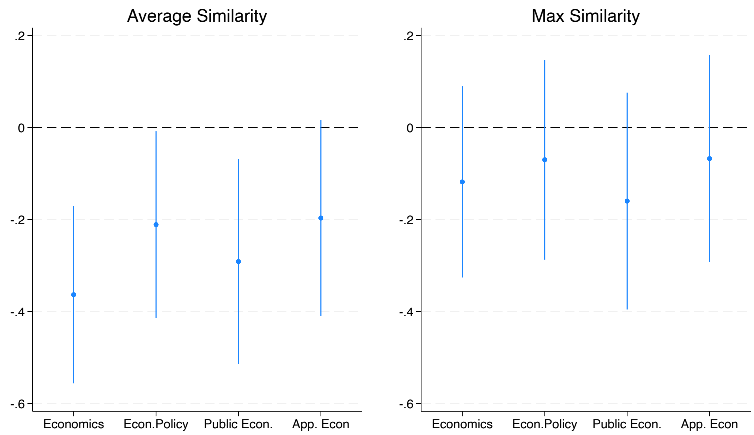
[30]We check the robustness of these results to using the continuous variable for similarity instead of the dummy variable in Table A.9. In line with Table 11, the coefficient of the female dummy is close to 0 and not statistically significant in almost all columns of Panel 1, while it increases in size and becomes highly statistically significant in columns 4-5 of Panel 2.

Table 11: Gender differences in similarity

	Mean	Mean	Mean	Max	Max	Max
	(1)	(2)	(3)	(4)	(5)	(6)
Panel 1: With female members of the committees only						
Female	-0.004	0.001	0.029	0.022	0.019	0.031
	(0.025)	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
Observations	1,782	1,782	1,782	1,782	1,782	1,782
R-squared	0.030	0.057	0.244	0.057	0.084	0.273
Panel 2: With male members of the committees only						
Female	-0.027	-0.029	-0.011	-0.043**	-0.041*	-0.014
	(0.022)	(0.021)	(0.022)	(0.022)	(0.021)	(0.021)
Observations	2,328	2,328	2,328	2,328	2,328	2,328
R-squared	0.031	0.072	0.201	0.055	0.087	0.278
$\overline{Y}$	0.5	0.5	0.5	0.5	0.5	0.5
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Call FE	No	No	Yes	No	No	Yes
Year FE	No	Yes	No	No	Yes	No
Broad Field FE	No	Yes	No	No	Yes	No

Notes. Dependent variables: Mean/Maximum Similarity between the candidate and female/male members of the selection committees. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 5: Gender differences in similarity by broad fields



Notes. The figure shows the coefficients of the interaction terms between the female dummy and the dummies for broad fields of the call (Equation 2). Econometrics is the omitted broad field.

6.3 Research similarity and the gender gap in the probability of winning

Having shown that there is a gender gap in (maximum) research similarity, we explore its role in explaining the presence of women in academia. To do this, we replicate the analysis in Table 3, which includes the probability of attending the interview in the controls, and add the mean or maximum similarity dummy to the regression. In other words, we add to equation 1 the estimate of a candidate's probability of attending the interview and explore the association between research similarity and the probability of winning.

The results are shown in Table 12 and confirm that research similarity is positively and significantly related to the probability of winning, both when we use the mean and the maximum similarity dummy and in all specifications considered. When we control for the probability of attending the interview, we observe that the coefficient of the female dummy is negative and significant in all specifications, implying a gender gap in the probability of winning. Interestingly, the gender gap in the probability of winning is generally smaller in magnitude compared to Table 3, especially in columns 4 and 5, where we use the maximum similarity dummy. This is consistent with the result that there are gender differences in research similarity only when the latter is measured by the maximum similarity with the committee members. According to these results, the inclusion of the maximum similarity dummy explains 3-4% of the gender gap reported in Table 3.[31]

[31] In Table A.10, we check that these results are robust to using the continuous version of our similarity variable.

Table 12: The role of research similarity in the probability of winning, LPM

Controlling for the probability of taking part to the interview

VARIABLES	(1) Winner	(2) Winner	(3) Winner	(4) Winner	(5) Winner	(6) Winner
Dummy Similarity	0.057***	0.053***	0.048***	0.054***	0.055***	0.046***
	(0.012)	(0.013)	(0.016)	(0.012)	(0.013)	(0.014)
Female	-0.075***	-0.071***		-0.074***	-0.069***	
	(0.016)	(0.017)		(0.016)	(0.017)	
PhD Abroad	0.009	0.008		0.009	0.007	
	(0.016)	(0.017)		(0.016)	(0.016)	
Abroad	-0.074***	-0.052***	0.012	-0.070***	-0.047***	0.013
	(0.016)	(0.017)	(0.020)	(0.016)	(0.017)	(0.019)
Years from PhD	-0.004*	-0.006***	0.065*	-0.004*	-0.006***	0.066*
	(0.002)	(0.002)	(0.038)	(0.002)	(0.002)	(0.038)
Internal Candidate	0.015	0.011	0.096**	0.016	0.009	0.097**
	(0.044)	(0.045)	(0.045)	(0.044)	(0.045)	(0.045)
Co-author	0.070	0.050	0.155**	0.060	0.034	0.145**
	(0.062)	(0.064)	(0.066)	(0.063)	(0.065)	(0.066)
At least one Top 6	-0.056	-0.085	-0.071*	-0.053	-0.088	-0.072*
	(0.067)	(0.074)	(0.038)	(0.066)	(0.075)	(0.038)
N pubs in A+	0.057***	0.067***	0.128**	0.058***	0.069***	0.129**
	(0.016)	(0.015)	(0.063)	(0.015)	(0.015)	(0.063)
At least one Interd.	-0.022	-0.017	-0.130***	-0.026	-0.021	-0.094*
	(0.043)	(0.048)	(0.050)	(0.043)	(0.048)	(0.052)
N pubs in A	0.026***	0.027***	0.014	0.026***	0.027***	0.012
	(0.003)	(0.004)	(0.012)	(0.003)	(0.004)	(0.012)
N pubs	-0.006***	-0.007***	-0.006	-0.007***	-0.008***	-0.005
	(0.001)	(0.001)	(0.009)	(0.001)	(0.001)	(0.009)
Pr (Present)	0.969***	1.026***	0.009	0.991***	1.055***	0.041
	(0.147)	(0.161)	(0.128)	(0.147)	(0.161)	(0.127)
	Mean	**Mean**	**Mean**	**Max**	**Max**	**Max**
$\overline{Y}$	0.10	0.10	0.10	0.10	0.10	0.10
Observations	2,365	2,365	2,365	2,365	2,365	2,365
R-squared	0.096	0.174	0.450	0.095	0.174	0.450
Call FE	No	Yes	No	No	Yes	No
Year FE	Yes	No	Yes	Yes	No	Yes
Broad Field FE	Yes	No	No	Yes	No	No
Candidate FE	No	No	Yes	No	No	Yes

Notes. Dependent variable: Probability of winning a senior assistant professorship. Estimates from a Linear Probability Model. Our controls include a dummy for candidates holding a PhD from abroad and for being currently abroad, number of years from PhD, a dummy for internal candidates and for co-authorship, a dummy for those with at least one Top 6 publication, number of A+ and A publications, number of publications, dummy for those with at least one publication in interdisciplinary journals. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6.4 Discussion

We interpret the premium that research similarity grants to candidates in the selection process as evidence that senior academics/evaluators rate junior researchers with research agendas similar to their own more positively. While there is evidence of gender homophily in Economics (Ductor and Prummer, 2024), self-image bias relates indirectly to gender through research agendas, which are the focus of this paper. In this section, we further discuss our results and alternative explanations for our findings.

It can be argued that selection committee members do not choose the winning candidate according to their own preferences, which may be influenced by self-image bias, but rather act on input from departments that want to hire junior researchers with specific research characteristics and identify members of selection committees with this goal in mind. We note that our results hold also when we only consider the external members of the committee and exclude the internal one, who represents the direct interest of the hiring department, as shown in Table 7. We have also shown that our results hold when controlling for the primary field of research of the internal member of the selection committee, which can be a proxy for the finer research field the department is targeting in the call (Table 5). In addition, similarity to the department has no positive effect on the probability of winning, when we control for similarity with the selection committee (Table 6, Panel 3). Moreover, in Figure A.13 we show that, controlling for the broad field of the call, the distribution of the mean and maximum similarity for members of the same selection committee are to the right and statistically different from the distribution of the mean and maximum similarity for members of different committees.[32] This suggests that selection committees are homogeneous groups in terms of research interests. On the one hand, this may indicate that departments have a particular research profile in mind. On the other hand, homogeneous research characteristics in the selection committee may signal that departments expect them to act according to a self-image bias and support a candidate similar to themselves rather than the most qualified one. For example, the evidence in Table 6 suggests that, in many specifications, similarity is more strongly related to the probability of winning than high ranked publications. We also point out that similarity plays an important role even controlling for the primary field of

[32]Specifically, according to the Kolmogorov-Smirnov test, the distributions for the mean similarity are different for Economics, Economic Policy and Applied Economics, and Econometrics, while the distributions for the max similarity are different for Economics, Economic Policy and Applied Economics.

research of the candidate, as shown in Table 6, Panel 2 and 3. It may also be that our results are not explained by self-image bias, but rather by a comparative advantage that committee members have in judging candidates more similar to themselves. However, we find no evidence that they are better at inferring the impact of the candidate's work. First, the positive influence of similarity is also present when we control for the citations of the candidate's publications (see Section 6.1.1 and Table 6). Moreover, if we test whether winning candidates with a higher similarity index are more productive after promotion, we do not find evidence confirming this hypothesis. The results are reported in Table A.11. As proxy of quality we use the publication record (i.e., average number of citations per publication, a dummy for A+ publications and a dummy for Top 6 publications) starting from the year after the call. In Panel 1, we focus only on winning candidates and look at the relationship between mean (or max) similarity of the winning candidate with the selection committee and their publication record after becoming senior assistant professors. In Panel 2, we extend the analysis to the entire sample of candidates and interact the dummy for the winning candidate with a variable measuring the (mean and max) similarity index of the winner applied to all candidates participating to the call. We find no evidence that higher similarity is associated with hirings of higher quality. Winning candidates with higher similarity index are not more productive after promotion compared to other winning candidates with a lower index. Moreover, the difference in post-productivity between the winning candidate and the other candidates participating to the same call is not larger in calls where the winning candidate has a higher similarity index.

An alternative explanation for our findings may be that candidates choose to participate in calls where they see that selection committee members have an agenda similar to their own. However, candidates do not know the composition of the selection committee at the time of the application, as we discussed in Section 2. Moreover, the evidence in Table A.6 (Panel 2) indicates that similarity is not related to the decision to participate to the interview, when shortlisted. Table A.8 explores the application margin using the pseudo panel described in Section 5.2. We find that similarity is not associated with the decision to apply to a specific call, with the exception of the max similarity index for small departments (Panel 1). This may indicate that in calls initiated by small departments, candidates tend to apply only if their research closely aligns with the interests of the department members. Furthermore, we find no evidence that the effect varies by gender, supporting the hypothesis that female and male candidates do not self-select differently into calls based on their level of similarity with the committee (Panel 2). Note also, that the pseudo panel may not reflect the universe of candidates who could have applied but did not, making this analysis tentative.

Finally, there may be a concern that our similarity measures capture closeness in language rather than in research characteristics, as discussed in Section 4.2. However, our model has consistently proven to be the most accurate for semantic text similarity. Although we cannot fully exclude that language may play a role, we note that more than 90% of our sample of candidates is made up of non-native English speakers. Therefore, the variation in our index cannot be driven by language differences between native and non-native English speakers. In addition, we focus on the abstracts of scientific publications (within fields), where differences in language use across genders are likely to be more limited. Finally, when we look at the distribution of our similarity indices by field, we do not find that such distributions in a field like econometrics, where the language is likely to be more standardised, largely differ from the distributions in other fields (see Figure A.14, where we distinguish by broad field of the call).

Overall, our evidence is consistent with the importance of similarity being driven by the demand side of the academic market, rather than the supply side, and with self-image bias playing a role in it. Since female candidates are less likely to be very similar to male members of selection committees, and since men make up a higher proportion of selection committee members, self-image bias may contribute to explaining part of the gender gap in economics.

7 Conclusions

There is extensive evidence of the (economic) benefits of a diverse and inclusive workforce. While measuring the benefits of diversity in the academic market can be challenging, studies show that it does impact on scholars' performance in measurable ways, such as citation counts (Powell, 2018). Diversity also enriches the scientific process by bringing in a wider range of perspectives and research questions. Promoting diversity in academia is therefore not only a matter of fairness, but also of efficiency.

In this paper, we analyse the presence of self-image bias in academia, which may play a role in the slow changes in gender diversity among scholars and in the narrowing of research agendas. We propose a novel and granular measure of similarity that captures not only research areas, but also broader characteristics of research agendas, starting from the abstracts of the papers. This new measure of similarity has the potential to capture the diversity of knowledge production better than fields of research, and to reveal research directions over time and space more accurately. We use it to investigate whether the similarity between selection committee members and candidates for senior assistant professorships is associated with the outcome of the selection process and whether female candidates are characterised by a lower similarity index with more senior academics than their male counterparts, offering an explanation for the gender gap in the probability of becoming a senior assistant professor.

In order to answer the research questions, we use data on the Italian academic job market, and collect the publications of the universe of candidates, members of recruitment committees and of faculty of departments opening calls for the period 2014-2021. By applying NLP techniques, we calculate an index of similarity between the publications of the committee members and those of the candidates, and show that candidates with a mean or maximum similarity larger than the median are between 4.5 and 6 percentage points more likely to win the competition for senior assistant professorships. The positive association between research similarity and the probability of winning is robust to a number of checks. We also find that women are, on average, less likely to be very similar to one of the committee members. This gender gap in maximum similarity is found only when looking at male committee members, while it disappears when we focus only on female members, and it may help explain the gender gap in the probability of winning that we observe once we control for selection into the interview.

The evidence presented suggests that in male-dominated contexts, similarity bias and the search for "fit" can hinder the career progression of female academics. Note that addressing similarity bias and promoting gender diversity in academia would not imply narrowing the topics researched in a department. The distribution of our similarity indices between candidates by gender suggests that the range of topics is the same for female and male candidates (Figure A.11). Thus, addressing self-image bias may, on the contrary, help to mitigate the tendency to conform to a standardised research profile, as has been observed in economics departments in recent years (Corsi *et al.*, 2019).

On the policy side, the identification of this source of bias provides additional justification for the implementation of affirmative action measures that deliberately increase the representation of minorities in the profession and, consequently, on selection committees. It also highlights the importance of establishing transparent and objective criteria in the evaluation process, while ensuring a degree of diversity in research interests among committee members. Failure to address self-image bias may perpetuate the gender imbalance in economics and limit innovative research.

Supplementary data

The data and codes for this paper are available on the Journal repository. They were checked for their ability to reproduce the results presented in the paper. The authors were granted an exemption to publish parts of their data because access to these data is restricted. However, the authors provided the Journal with temporary access to the data, which enabled the Journal to run their codes. The codes for the parts subject to exemption are also available on the Journal repository. The restricted access data and these codes were also checked for their ability to reproduce the results presented in the paper. The replication package for this paper is available at the following address: https://doi.org/10.5281/zenodo.17397753.

Affiliations

Piera Bello, Department of Economics, University of Bergamo, piera.bello@unibg.it

Alessandra Casarico, Department of Social and Political Sciences, Bocconi University, Milan; Dondena Research Center, Milan; CESIfo, Munich alessandra.casarico@unibocconi.it;

Debora Nozza, Department of Computing Sciences, Bocconi University, Milan, debora.nozza@unibocconi.it.

References

Ash, E., Asher, S., Bhowmick, A., Bhupatiraju, S., Chen, D., Devi, T., Goessmann, C., Novosad, P. and Siddiqi, B. (2023). 'In-group bias in the Indian judiciary: Evidence from 5 million criminal cases', Center for Global Development.

Ash, E., Chen, D.L. and Ornaghi, A. (2024). 'Gender attitudes in the judiciary: Evidence from US circuit courts', *American Economic Journal: Applied Economics*, vol. 16(1), pp. 314–50, doi:10.1257/app.20210435.

Auriol, E., Friebel, G., Weinberger, A. and Wilhelm, S. (2022). 'Underrepresentation of women in the economics profession more pronounced in the United States compared to heterogeneous Europe', *Proceedings of the National Academy of Sciences of the United States of America*, vol. 119(16), p. e2118853119, doi: 10.1073/pnas.2118853119.

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

Baltrunaite, A., Casarico, A. and Rizzica, L. (2023). 'Women in economics: The role of gendered references at entry in the profession', *Centre for Economic Policy Research Discussion Paper*, (17474).

Bayer, A. and Rouse, C.E. (2016). 'Diversity in the economics profession: A new attack on an old problem', *Journal of Economic Perspectives*, vol. 30(4), pp. 221–242.

Belot, M., Kurmangaliyeva, M. and Reuter, J. (2023). 'Gender diversity and diversity of ideas', Institute of Labor Economics (IZA).

Beneito, P., Boscá, J.E., Ferri, J. and García, M. (2021). 'Gender imbalance across subfields in economics: when does it start?', *Journal of Human Capital*, vol. 15(3), pp. 469–511.

Biasi, B. and Ma, S. (2022). 'The education-innovation gap', National Bureau of Economic Research.

Card, D. and DellaVigna, S. (2013). 'Nine facts about top journals in economics', *Journal of Economic literature*, vol. 51(1), pp. 144–161.

Chari, A. and Goldsmith-Pinkham, P. (2017). 'Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute', National Bureau of Economic Research, Inc.

Chen, Y., Mahoney, C., Grasso, I., Wali, E., Matthews, A., Middleton, T., Njie, M. and Matthews, J. (2021). 'Gender bias and under-representation in natural language processing across human languages', in (*Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*pp. 24–34.

Conneau, A. and Kiela, D. (2018). 'Senteval: An evaluation toolkit for universal sentence representations', in (*Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*European Language Resources Association (ELRA).

Corsi, M., D'Ippoliti, C. and Zacchia, G. (2019). 'Diversity of backgrounds and ideas: The case of research evaluation in economics', *Research Policy*, vol. 48(9), p. 103820.

De Paola, M. and Scoppa, V. (2015). 'Gender discrimination and evaluators' gender: Evidence from italian academia', *Economica*, vol. 82(325), pp. 162–188.

Ductor, L. and Prummer, A. (2024). 'Gender homophily, collaboration, and output', *Journal of Economic Behavior & Organization*, vol. 221(C), pp. 477–492.

Dupas, P., Modestino, A.S., Niederle, M., Wolfers, J. and the Seminar Dynamics Collective (2021). 'Gender and the dynamics of economics seminars', National Bureau of Economic Research.

Eberhardt, M., Facchini, G. and Rueda, V. (2023). 'Gender differences in reference letters: Evidence from the economics job market', *The Economic Journal*, vol. 133(655), pp. 2676–2708, doi:10.1093/ej/uead047.

Elsevier (2023). 'Scopus. abstract and citation database. www.scopus.com', .

Gneezy, U. and Rustichini, A. (2004). 'Gender and competition at a young age', *American Economic Review*, vol. 94(2), pp. 377–381.

Hengel, E. (2022). 'Publishing while female: Are women held to higher standards? evidence from peer review', *The Economic Journal*, vol. 132(648), pp. 2951–2991.

Hill, T., Smith, N.D. and Hoffman, H. (1988). 'Self-image bias and the perception of other persons' skills', *European Journal of Social Psychology*, vol. 18(3), pp. 293–298.

Huber, L., Memmadi, C., Dargnat, M. and Toussaint, Y. (2020). 'Do sentence embeddings capture discourse properties of sentences from scientific abstracts?', in (*Proceedings of the First Workshop on Computational Approaches to Discourse*pp. 86–95, Association for Computational Linguistics, doi:10.18653/v1/2020.codi-1.9.

Koffi, M. (2021). 'Gendered citations at top economic journals', in (*AEA Papers and Proceedings*pp. 60–64, vol. 111.

Koffi, M. (forthcoming). 'Innovative ideas and gender inequality', *American Economic Review*.

Le Barbanchon, T., Rathelot, R. and Roulet, A. (2021). 'Gender differences in job search: Trading off commute against wage', *The Quarterly Journal of Economics*, vol. 136(1), pp. 381–426.

Lundberg, S. and Stearns, J. (2019). 'Women in economics: Stalled progress', *Journal of Economic Perspectives*, vol. 33(1), pp. 3–22.

Niederle, M. and Vesterlund, L. (2007). 'Do women shy away from competition? Do men compete too much?', *The Quarterly Journal of Economics*, vol. 122(3), pp. 1067–1101.

Paredes, V., Paserman, M.D. and Pino, F.J. (2023). 'Does economics make you sexist?', *The Review of Economics and Statistics*, vol. 105, pp. 1–47.

Powell, K. (2018). 'These labs are remarkably diverse–here's why they're winning at science', *Nature*, vol. 558(7708), pp. 19–23.

Reimers, N. and Gurevych, I. (2019). 'Sentence-BERT: Sentence embeddings using siamese BERT-networks', in (*Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*pp. 3982–3992, Hong Kong, China: Association for Computational Linguistics, doi:10.18653/v1/D19-1410.

Sarsons, H. (2017). ‘Recognition for group work: Gender differences in academia’, *American Economic Review*, vol. 107(5), pp. 141–45.

Sierminska, E. and Oaxaca, R.L. (2021). ‘Field specializations among beginning economists: Are there gender differences?’, in (*AEA Papers and Proceedings*pp. 86–91, vol. 111.

Siniscalchi, M. and Veronesi, P. (2021). ‘Self-image bias and lost talent’, Becker Friedman Institute for Economics, University of Chicago, doi:10.2139/ssrn.3820954.

Stansbury, A. and Schultz, R. (2023). ‘The economics profession’s socioeconomic diversity problem’, *Journal of Economic Perspectives*, vol. 37(4), pp. 207–230.