BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

Women in economics: the role of gendered references at entry in the profession

by Audinga Baltrunaite, Alessandra Casarico and Lucia Rizzica

February 2024

Number 1438

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working Papers)

Women in economics: the role of gendered references at entry in the profession

by Audinga Baltrunaite, Alessandra Casarico and Lucia Rizzica

Number 1438 - February 2024

The papers published in the Temi di discussione *series describe preliminary results and are made available to the public to encourage discussion and elicit comments.*

The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank.

ISSN 2281-3950 (online)

Designed by the Printing and Publishing Division of the Bank of Italy

WOMEN IN ECONOMICS: THE ROLE OF GENDERED REFERENCES AT ENTRY IN THE PROFESSION

by Audinga Baltrunaite*, Alessandra Casarico** and Lucia Rizzica***

Abstract

We study the presence and the extent of gender differences in reference letters for graduate students in economics and finance, and how these differences relate to early labor market outcomes. To these ends, we build a novel rich dataset and combine Natural Language Processing techniques with standard regression analysis. We find that men are described more often as brilliant and women as hardworking and diligent. We show that the former (latter) description relates positively (negatively) with various subsequent career outcomes. We provide evidence that the observed differences in the way candidates are described are driven by implicit gender stereotypes.

JEL Classification: I23, J16, J44.
Keywords: gender bias, research institutions, professional labor markets, word embeddings.
DOI: 10.32057/0.TD.2023.1438

* Bank of Italy and CEPR.

** Bocconi University, CESIfo and Dondena.

*** Bank of Italy.

1 Introduction[*]

The under-representation of women in academic ranks is a widespread phenomenon and it has barely changed over time, especially in some fields of study. As discussed by Bayer and Rouse (2016) and Lundberg and Stearns (2019), within the field of economics, women are a minority starting from the undergraduate level, and this gap widens when looking into the higher ranks of academia. Such gender imbalance "likely hampers the discipline, constraining the range of issues addressed and limiting the ability to understand familiar issues from new and innovative perspectives" (Bayer and Rouse, 2016, p.221).

The leaky pipeline phenomenon is by no means limited to economics, as testified by the evidence collected in the work of the European Commission (2021), and it is particularly severe in STEM disciplines (Kim and Moser, 2021). Research on the causes of the leaky pipeline in academia, as well as of the low female presence in key and influential institutions, has developed a lot in the last decade, prompted by an increased awareness of the costs for the profession and for society at large of such under-representation. Lundberg (2020) takes stock of the research on women in economics and provides a comprehensive overview of the available evidence and explanations on why women are still a minority in the field, following all the stages of the career. The research reviewed in her work highlights several sensitive stages of the career: major choice at the undergraduate level, entry and performance in graduate school, publication records, maternity.

In this paper we focus on the transition from postgraduate program to work, and contribute to the literature by delving into how implicit attitudes held by senior academics relate to the under-representation of women in economics by influencing the early stages of

[*] We thank Francesco Billari, Claudia Ciccodicola, Mariateresa Elia, Francesca Felice, Sabina Marchetti, Gemma Prunner-Thomas, Riccardo Russo, Giovanni Salvucci, Nicola Scalzo, and Ivan Triglia for their invaluable help with data acquisition and analysis. Moreover, we are indebted to Jonathan Rossi, Giorgia Cascone, Alessandro Izzo, Madeline Kett, Nina Lurati, Julia Milis, Marvin Pappalettera, and Luigi Roncoroni for excellent research assistance and we are grateful to Elliott Ash, Sonia Bhalotra, Luca Bellodi, Michela Carlana, Marta De Philippis, James Fenske, Sergio Galletta, Silvia Giacomelli, Stephen Hansen, Dirk Hovy, Eliana La Ferrara, Tommaso Orlando, Francesco Principe, Jesper Roine, Pilar Sorribas Navarro, and seminar participants at Bank of Italy, Bocconi University, University College London, Tinbergen Institute, Free University Berlin, COSME Gender Economics Workshop, 4th Monash-Warwick-Zurich Text-as-Data Workshop, University of Bergamo, Catholic University-Milan, University of Urbino, DG Ecofin, FROGEE-FREE Workshop in Warsaw, and Royal Economic Society Conference for insightful comments. *The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy and of Bocconi University. The usual disclaimers apply.*

the career. Specifically, we examine whether advisors write reference letters differently for male and female PhD candidates, conditional on observed candidate quality. We focus on the content of the letters that PhD candidates receive when applying to assistant professor (or equivalent) positions in the economics job market and, by combining modern text analysis tools with insights from the psychology literature, explore whether such letters reveal implicit gender stereotypes in how candidates are talked about and who holds such stereotypes. Finally, and importantly, we estimate how candidate, advisor and letter characteristics, and implicit biases the letters may contain relate to early career outcomes.

In order to conduct the analysis, we collect data from two large institutions – one in academia and one in the government sector – recruiting internationally on the academic job market for junior economists. They are both based in Italy. We gather information on ten years of applications, for a total of about 8,000 applications and 25,000 reference letters. We recover detailed information on applicants from their application packages and conduct text analysis on their reference letters employing Natural Language Processing (NLP) tools (specifically, word embeddings). Finally, we map candidates to their position (i.e., placement institution and position in the academic job ladder) and publication records using massive web-scraping techniques on several publicly accessible websites.

We find that there are significant differences between male and female candidates when they get to the job market. First, female candidates are more likely to have female advisors, who are generally more junior and less established in the academic community. Second, they receive fewer letters, both because they have fewer sponsors and because there is a higher incidence of advisors not submitting the letter when they are supposed to. Finally, and most importantly, the content of the letters is consistently different: letters written for female candidates tend to stress more their hardworkingness and diligence, rather than their brilliance and smartness. Regressing candidate early career success proxies – rank of the placement institution, position in the academic ladder, number and quality of publications and citations – on the characteristics of reference letters, we find that how candidates are talked about significantly relates to candidates' success in the economics profession. Finally, we present evidence to try to disentangle the underlying mechanisms. First, we argue that these differences in candidates' characterizations are not explained by intrinsic differences between male and female candidates. Indeed, we show that only male advisors tend to talk about male and female candidates differently. Second, we exclude that these differences in candidate descriptions are demanded by the market by investigating gender-specific returns to grindstone and standout characterizations. The finding that only men benefit from being described as standout, while being presented as grindstone is detrimental to both men and women, points against this demand-side explanation. We thus argue that gender differences in candidates' characterization stem from implicit biases that senior professionals in economics may hold.

Our findings are relevant from different standpoints. While outright discrimination is harder to go undetected now compared to the past, implicit bias may be persistent and difficult to capture. Using word embeddings, we aim to provide evidence on the presence of such implicit stereotypes in a natural, rather than experimental (Carlana, 2019), setting. Our analysis can advance our knowledge on the roots of the under-representation of women in academic and institutional ranks by opening up the black-box of gendered mentorship configurations and by studying how language used in the reference process can vehicle implicit biases and, through these, influence career outcomes. The link between letter characteristics and how candidates are characterized in those letters and their subsequent labor market outcomes is indeed a key novel aspect of our analysis. Eventually, we are going to shed light on the (potential) presence of "institutional discrimination" – i.e., when the rules, practices, or non-conscious understandings of appropriate conduct systematically advantage or disadvantage members of particular groups (Haney-Lopez, 2000) – in the academic job market process and, more in general, in all referral-based career mechanisms.

The remainder of the paper is structured as follows: in Section 2 we review the literature that is most closely related to our work; in Section 3 we introduce our data collection process and provide some descriptive statistics; Section 4 is dedicated to the text analysis of the reference letters; in Section 5 we present our estimation strategy, show the results of the regression analysis and provide robustness checks. Section 6 discusses the channels behind observed gender differences in candidate characterization in reference letters. Section 7 concludes.

2 Related literature

This paper stands at the junction between two fields of study: on the one hand, the economics literature that has started digging into the roots of the observed gender imbalance in the profession; on the other hand, a more established literature in applied psychology that aims to pin down the presence and magnitude of stereotypes and implicit (gender) bias in the labor market. Drawing from the latter to qualify the relevant stereotypes, we employ the tools of modern text analysis to quantify whether such stereotypes appear in a large corpus of reference letters and how they affect women's careers.

Within the field of economics, the literature has extensively documented the gender divide in academia (Bayer and Rouse, 2016; Janys, 2022). In US top departments women represent 18% of full and 26.5% of associate professors (Chari, 2023). In Europe, the share of women working in academic departments is overall 32%, and it becomes 27% in senior positions (Auriol et al., 2022). These imbalances appear very early in the career and have changed very little in recent years. According to Lundberg (2020), in the US the share of women among assistant professors in top departments has stalled over the last decade, while that of women in more senior positions has increased slowly.

A growing literature has highlighted different mechanisms contributing to the observed under-representation of junior female professionals in economics.

Focusing on the graduate level stage, some factors that are positively correlated with female PhD success include hiring and retaining female faculty, requiring student work-in-progress seminars, a more supportive seminar culture, and general awareness of gender bias issues (Boustan et al., 2020). The gender mix of peers in doctoral programs is also important: a higher fraction of women in entering PhD cohorts would reduce the gender gap in program attrition, with the effect driven almost entirely by differences in the probability of dropping out in the first year of the program (Bostwick and Weinberg, 2020).

The matching of students and advisors by gender and how such pairing affects job market outcomes are analyzed in Neumark and Gardecki (1998) and Hilmer and Hilmer (2007). The first survey several cohorts of graduate students from institutions granting PhDs in economics from the mid 1970s until the early 1990s and find no link between gendered student-advisor matching and rank of institution of placement. They do find evidence, though, that female students complete their PhD more often and more quickly when paired with a female advisor. Hilmer and Hilmer (2007), instead, focus on 1,900 individuals receiving economics PhDs from the top-30 Economics programs between 1990 and 1994 and examine the differential impact of each of the four possible mentorship configurations (female student–female advisor, female student–male advisor, male student–female advisor, and male student–male advisor) on both initial job placements and early-career research productivity, finding that the female–female pairing is worse than the male–male one, but no worse than the female-male.[1] Finally, Pezzoni et al. (2016) look at the impact of gender pairing of advisors and their students on research performance of Caltech students during graduate studies. Their evidence suggests that both male and female students publish more when paired with female advisors.

Another important factor affecting PhDs' placement is the field of specialization. Fortin et al. (2021) show that the gap in the likelihood of obtaining an assistant professor position in an institution outside the Top-50 can be fully explained by differences in the field of specialization; in the Top-50 departments, instead, the institution granting the PhD is the most powerful predictor. Similarly, looking directly at earnings, Oaxaca and Sierminska (2021) conclude that 14 percent of academics of either sex would have to change specialization in order to achieve complete salary parity across genders. Recent work by Belot et al. (2023) shows that differences in placement success can be explained by "idea homophily",[2] such that senior academics tend to systematically prefer candidates with a research agenda more similar to theirs. This implies that female PhDs specializing in highly male dominated fields fare significantly better than those in traditionally female dominated fields.

Some very recent literature, moreover, has started highlighting the existence of non observable obstacles and implicit discrimination specifically in the field of economics. Paredes et al. (2023), for instance, provide evidence that implicit and explicit gender stereotypes are well present in economics from the undergraduate level, with students turning out to be more gender biased than those in other fields and with the gap increasing over the course of studies. Looking directly at faculty members and exploiting the introduction of blind grading of exams in economics at Stockholm University, Jansson and Tyrefors (2022) show that teachers tend to give higher grades to male students.[3]

[1] Focusing on chemistry, Gaulé and Piacentini (2018) find that students working with advisors of the same gender tend to be more productive during the PhD, and that female students working with female advisors are considerably more likely to become faculty themselves. Hence, they argue that the under-representation of women in science and engineering faculty positions may perpetuate itself through the lower availability of same-gender advisors for female students.

[2] Bello et al. (2023) provide similar evidence at a later stage of the career, i.e., at the moment of tenure track application.

[3] Evidence on teachers' gender stereotypes is provided by Carlana (2019), for primary schools, and by Bleemer (2019) at the undergraduate level. The latter, in particular, focuses on the degree of "genderedness" of students' evaluations written by UC Santa Cruz professors and estimates the impact of such trait on the subsequent major choices by student.

Stark gender differences further appear in the process of publication of scientific work. Sarsons (2017) and Sarsons et al. (2021) show that women obtain less recognition for their contribution in coauthored research when collaborating with men, i.e., that coauthored papers affect the probability of being granted tenure less for female economists than for male economists. Similarly, Hengel (2022) employs several NLP tools to show that female authored papers are held to higher writing standards by editors and referees, so that women need to put significantly more effort for publishing their work, and hence are eventually less productive. Koffi (2021a,b) shows that female-authored papers published in the economics Top 5 journals are significantly less likely to be cited than those written by men, even when they are equally closely related to the research considered. However, Card et al. (2020) document no gender disparities along the whole publication process in top economic journals.

Finally, some recent papers have shown how male economists' attitudes expressed in public may be further detrimental to their female colleagues. Dupas et al. (2021) analyze interactions during seminars in economics departments to find that female presenters are systematically asked more and harsher questions by male audience. Sarsons and Xu (2021) survey male and female economists from top departments and show that men are systematically more self-confident than women, providing strong personal judgments even when asked questions on the economy, which are further away from their field of expertise.

Our paper contributes to this literature and sheds further light on the channels through which gender gaps are generated at the early stages of the economics profession. Specifically, we add to the existing contributions by studying the extent and impact of implicit gender stereotypes held by senior faculty.

We borrow from a literature coming from the fields of psychology and linguistics that has studied the presence of gender stereotyping in reference letters. Trix and Psenka (2003) analyze a corpus of about 300 letters of recommendation for medical faculty at a large American medical school and find that letters written for female applicants differ systematically from those written for male applicants in the extremes of length, in the percentages without basic features, and in the percentages with doubt raisers. Dutt et al. (2016) focus on geo-science and examine the relationship between applicant gender and two outcomes: letter length and letter tone. They show that female applicants are only half as likely to receive excellent letters versus good letters compared to male applicants. In addition, male and female recommenders differ in their likelihood to write stronger letters for male applicants over female applicants.

Some works, then, have specifically investigated the content of reference letters in various contexts, using pre-defined semantic classifications. Schmader et al. (2007) examine a corpus of reference letters written for applicants for either a chemistry or biochemistry faculty position at a large U.S. research university. Their findings, though based on a fairly limited sample - 886 letters of recommendation written on behalf of 235 male and 42 female candidates - reveal that recommenders tend to use significantly more "standout" terms to describe male as compared to female candidates. Letters containing more standout words also include fewer "grindstone" words.[4] In a similar spirit, Madera et al. (2009) analyze letters written for applicants to faculty positions in a psychology department, searching for descriptions of candidates that reflect a social role theory of sex differences.[5] The authors find that women are indeed described as more communal and less agentic than men, and that communal characteristics negatively affect the hiring decisions. The latter finding is based on judgments of hireability made by psychologists. Finally, Chapman et al. (2020) carry out a comprehensive study of letters of recommendation for a pool of Radiation Oncology Residency Applicants. Similarly to the previously mentioned studies, they use a dictionary of predetermined themes (LIWC) including standout, grindstone, agentic, communal and also other personality traits. While they do not detect significant differences depending on the gender of the applicant, they document significant linguistic differences related to the gender and other characteristics of the letter-writer, with a general tendency to use a male-biased language.[6] Importantly, in spite of the evidence on how men and women are described in

[4]Standout words include those referring to the exceptional characteristics of the person or item described. These include, for example, "outstanding", "exceptional", "unique", etc. Grindstone words instead refer to the effort a person exerts in her work. These include for example "hardworking", "tenacious", "work ethic".

[5]According to social role theory (Eagly et al., 2000), behavioral sex differences arise from the division of labor—the differential social roles inhabited by women and men. Historically, men have been more likely to engage in tasks that require speed, strength, and the ability to be away from home for expanded periods of time, whereas women were more likely to stay home and engage in family tasks, such as child rearing. Accordingly, men are perceived and expected to be agentic, and women are perceived and expected to be communal. Agency includes descriptions of aggressiveness, assertiveness, independence, and self-confidence (Eagly and Johannesen-Schmidt, 2001). Agentic behaviors at work include speaking assertively, influencing others, and initiating tasks. Communal behaviors at work include being concerned with the welfare of others (i.e., descriptions of kindness, sympathy, sensitivity, and nurturance), helping others, accepting others' direction, and maintaining relationships (Eagly and Johannesen-Schmidt, 2001).

[6]Language is evaluated for gender bias using a publicly available gender bias calculator, available here.

reference letters in terms of standout, grindstone, communal or agentic words, the available evidence from the psychology literature does not support the view that men and women are intrinsically different with respect to these characteristics (Hyde, 2014).

These two strands of literature – that of the economics of gender gaps and that of linguistic analysis of reference letters – have evolved separately: there is no contribution jointly investigating the extent of the bias of sponsors or advisors and the impact this has on real labor market outcomes. Our paper aims to fill this gap by evaluating the extent of bias in reference letters and estimating how gender differences in the language used to describe candidates influence subsequent professional outcomes in a set up in which rich student and advisor characteristics are controlled for. While our analysis is guided by the evidence provided in the psychology literature mentioned above, we advance on these studies taking a massive data analysis approach: we examine around 25,000 letters using modern tools of text analysis (word embeddings) and then incorporate our innovative measures of implicit stereotypes in a novel and rich dataset, which covers graduate students' and advisors' characteristics, so as to estimate comprehensive regression models.

Contemporaneous work by Eberhardt et al. (2023) provides an analysis of economics job market reference letters along similar lines. The authors rely on a fairly small sample of reference letters, which only allows them to apply word count tools to extract candidate descriptions from the content of the text. Similar to our results, they signal a prevalence of grindstone terms in female candidates' description vis-à-vis a more pronounced use of standout words for male candidates. The former characterizations are shown to correlate with the ranking of the job market placement institution in a way that is detrimental to women. Our work advances on these findings by building a considerably larger and richer dataset, which allows us to employ more advanced NLP techniques, i.e., word embeddings. This methodology permits to represent letter language in a way that preserves semantic relationships in the text and thus is able to take into account more information compared to count-based approaches, that would likely omit implicit associations. In addition, we carry out a more comprehensive regression analysis and look at a wider range of candidates', letters', and referees' characteristics and of career outcomes. Finally, we provide an evidence-based discussion of mechanisms behind our results pointing to the presence of implicit gender biases among letter writers as the main driver of gender differences in candidates' descriptions.

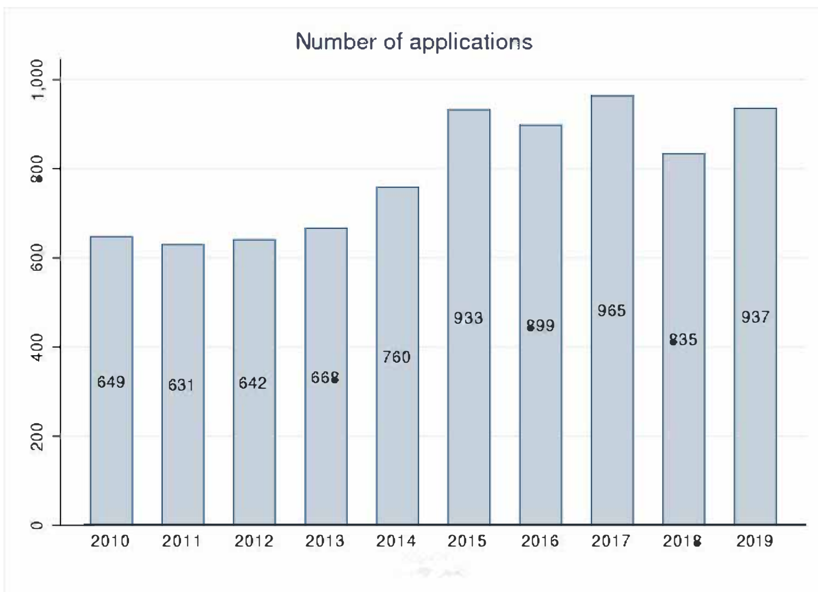
3 Data and descriptive evidence

3.1 Data sources

Our work draws from a novel unique dataset that we built for the project. Specifically, we collected the full package of applications received by two leading institutions recruiting on the international economics job market for assistant professor (or equivalent level) positions in Italy.[7]

The data cover ten cohorts of applicants for the academic institution – of which we have data for two departments (Economics and Finance) – and five cohorts for the other. Overall, we have data for almost 8,000 applications.[8] Figure 1 shows the distribution of the applications in our sample across years.

Figure 1: Number of applications by year of application.



Notes: Number of applications in the sample, by year. Years 2015-2019 include the two institutions, whereas years 2010-2014 only the academic one.

For each candidate we collect information at the time of application available in their CVs and application forms. These allow us to recover the institution in which they obtained

[7]Access to the data was granted under strict confidentiality agreements.

[8]Over 800 applications are repeated across institutions or departments. We drop duplicate observations. See Section 3.3 for more details.

their PhD, the main fields of interest[9] and some demographic and career information. We complement this information with that on their job market paper – i.e., whether and where it is published – to have a further proxy of candidate quality. We infer the gender of the applicant through gender name libraries and, in cases in which gender assignment is uncertain, we proceeded with manual checking. We then match the institution awarding the PhD with the (yearly) QS world university rankings and with the (2021) Repec ranking of Economics departments to obtain a proxy of PhD quality.

Each candidate's application package also contains the identities of their advisors who are to send their reference letters. Candidates indicate from two to five letter writers, for a total of 25,778 references. For each reference we can classify the gender of the letter writer and her main affiliation. The actual number of letters in the application package sometimes is lower than the number indicated by the candidate in the application form. This happens when sponsors do not send their reference letter (in time) to the institution.

Finally, we collect information on candidates' labor market outcomes through web-scraping and manual search. First, we collect information on candidates' first placement within three years after the job market. To this end we draw from the online Scopus repository and from LinkedIn. From Scopus we retrieve the candidate's affiliation as indicated on the first publication registered within a three year period from the job market application. Searching LinkedIn, then, we are able to match the first job title indicated in the three years following the application. We complement this information by manually searching for the missing career histories in candidates' web pages (when available). We also collect detailed information on candidates' current placement by means of massive web-scraping of three publicly available websites: Repec, Google Scholar and LinkedIn.[10] The first three allow us to collect comprehensive information on candidates who pursued a career in academia and research. Specifically, we retrieve the number and full list of publications, the number of citations and the main current affiliation. The LinkedIn platform, instead, allows us to obtain information on those candidates who pursued a non-research career or have not published any work yet. All in all, the combination of these three sources allows us to identify the current placement of 94% of the candidates. Note that for more recent candidates, first

[9]This was provided by the candidate in an open-ended question. We thus categorized the answers into JEL codes.

[10]The measures of current placement refer to the placement as indicated in our online sources at the time we scraped the web, i.e., between October 2020 and April 2021.

and current affiliation may coincide. Finally, as we do for the institution granting the PhD, for those in academia, we further match their first and current affiliation with measures of academic ranking taken from both QS and Repec, to obtain a proxy of their success on the job market.

3.2 Descriptive statistics

We now present descriptive statistics on the sample of candidates and letter writers.

As Figure 2 shows, less than one third of applications come from female candidates, a share that has remained constant in the ten years considered (left panel). The share of letters written by female sponsors is significantly lower, equal at most to 15% (right panel) and barely rising over the period.

Figure 2: Applications and reference letters by gender and year of application.

Notes: Years 2015-2019 include the two institutions, whereas years 2010-2014 only one.

Table 1 summarizes the main characteristics of the job market candidates in our sample and examines the gender differences in their observable characteristics (that may proxy, at least in part, for candidate quality) at the onset of their job market search. Moreover, it also includes information on references, the job market paper, and the subsequent labor market outcomes. Overall, the external validity of our sample is rather large as it is representative of at least the pool of candidates on the European economics junior job market. Indeed, Appendix Table B.1 provides similar descriptive statistics for the sample of European Job market candidates that subscribed to the European Economic Association Candidate Directory for the year 2020/2021.

Around a half of all candidates apply with a PhD from an institution in the United States or Canada, with more males than females (53% vs. 48%) having studied in North America.[11] The opposite is true for European PhDs (43% vs. 48%). With respect to the location of positions advertised by the two institutions, we observe that the pool of applicants is largely international: only 7% of all applicants receive their degree from an institution in Italy, on average. "Domestic" PhD is more common among female candidates: one out of ten women hold or are expected to hold a PhD from an Italian institution. Overall, thus, female candidates tend to come from geographically closer institutions, perhaps signaling their lower willingness to relocate during the job market.[12]

There are significant gender differences in terms of field of specialization of PhD candidates. Female PhD candidates are 10 percentage points more likely to specialize in applied microeconomics research relative to their male peers, who instead tend to choose topics in macroeconomics, finance, theory or quantitative methods more often.[13] Figure 3 illustrates these differences more in detail, by focusing on 14 categories based on the main JEL codes. Gender differences are mostly driven by macroeconomics or mathematics and quantitative methods (more often chosen by male candidates) and labor economics, demography or development economics (more often chosen by female candidates). Interestingly, there are no significant gender differences in financial or international economics.

In terms of the quality of the PhD granting institution, the pool of male applicants appears to be better selected: they more often come from a Top 20 institution according to either QS or Repec rankings. Finally, male applicants report more publications in their CVs at the time of application: over 70% of male candidates have at least one publication of some kind, while this is the case for only 53% of females.

Significant gender differences are also visible in the application package of candidates.

[11] This may be different from the institution of affiliation at the time of the job market application for (a modest fraction of) candidates applying after the conferral of the degree, e.g., the ones applying from a postdoctoral program.

[12] More broadly, this is in line with the literature showing that women have a lower propensity to move away from home for work or study (Rizzica, 2013).

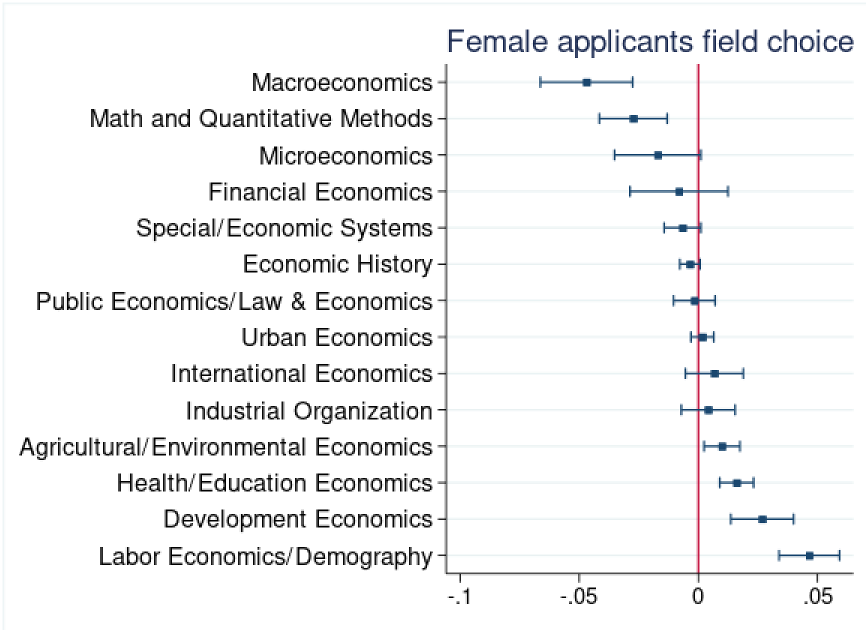
[13] Applied microeconomics includes JEL codes H, I, J, K, L, N, O, P, Q, and R. Macro/International/Finance include JEL codes E, F, G, and M. Theory/Quantitative covers JEL codes C and D.

Table 1: Descriptive statistics of job market candidates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Male		Female		Difference		
	mean	sd	mean	sd	N	Diff	T-stat
Pre-JM:							
US/Canada PhD	0.53	0.50	0.48	0.50	7077	0.048***	(3.647)
EU PhD	0.43	0.50	0.48	0.50	7077	-0.044***	(-3.329)
Italian PhD	0.07	0.25	0.10	0.30	7077	-0.033***	(-4.454)
Applied micro	0.24	0.43	0.34	0.48	7077	-0.100***	(-8.229)
Macro/International/Finance	0.44	0.50	0.40	0.49	7077	0.048***	(3.750)
Theory/Quantitative	0.24	0.43	0.20	0.40	7077	0.044***	(4.151)
Top 20 QS (general)	0.17	0.38	0.15	0.36	7063	0.020**	(2.143)
Top 20 Repec Econ	0.27	0.44	0.21	0.41	7063	0.052***	(4.725)
# Publication pre-JM	0.72	2.17	0.53	1.74	7077	0.186***	(3.784)
References:							
# Letter writers	3.25	0.78	3.21	0.84	7077	0.038*	(1.761)
# Letters uploaded	2.70	1.30	2.63	1.36	7077	0.079**	(2.250)
# Female letter writers	0.39	0.61	0.58	0.74	7077	-0.190***	(-10.320)
Main advisor female	0.11	0.31	0.17	0.37	6913	-0.055***	(-5.833)
Job Market Paper:							
Published job market paper	0.20	0.40	0.19	0.39	7077	0.013	(1.293)
Published job market paper in a Top 5	0.02	0.13	0.01	0.11	7077	0.004	(1.319)
Published job market paper in a Top 20	0.03	0.18	0.03	0.17	7077	0.006	(1.255)
Ranking of job market paper journal (Scimago 2021)	151.94	191.66	153.46	185.84	1156	-1.520	(-0.122)
# coauthors in job market paper	0.89	1.98	0.77	1.21	1387	0.127	(1.446)
Time to job market paper publication	2.01	6.09	2.40	5.95	1381	-0.390	(-1.080)
First placement:							
Academic placement Linkedin	0.81	0.39	0.81	0.39	5803	-0.002	(-0.168)
Placement Top 20 Repec Econ	0.11	0.31	0.11	0.31	5606	-0.001	(-0.083)
Assistant professor or higher	0.58	0.49	0.57	0.50	4401	0.010	(0.591)
PostDoc	0.22	0.41	0.25	0.43	4401	-0.025*	(-1.745)
Current placement:							
Academic placement Linkedin	0.75	0.43	0.75	0.43	6641	0.007	(0.597)
Placement Top 20 Repec Econ	0.08	0.26	0.08	0.27	6641	-0.006	(-0.807)
Associate professor	0.17	0.37	0.12	0.32	6641	0.050***	(5.498)
Assistant professor	0.46	0.50	0.50	0.50	6641	-0.036***	(-2.667)
PostDoc	0.12	0.32	0.13	0.34	6641	-0.014	(-1.495)
Research output:							
# Publications	2.37	5.49	1.54	3.66	7077	0.838***	(7.477)
Top 5 publication	0.07	0.25	0.05	0.21	7077	0.020***	(3.354)
# Citations (Repec)	41.07	147.87	26.18	78.62	7077	14.884***	(5.482)
Observations	5041		2036		7077		

Notes: All current placement and research output variables refer to 2021, while first placement variables refer to positions within three years after the job market application. * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

Figure 3: Differences in gender distribution across fields.



Notes: Each point is the difference between female and male candidates in the likelihood of declaring each field as main field (i.e., it is the coefficient associated to a female dummy variable in a linear regression with field dummies as outcomes).

Male candidates have a slightly higher number of academic references, both in terms of designated referees and of actual letters uploaded. Furthermore, there is evidence of assortative matching based on gender between students and advisors: female candidates have a higher number of female letter writers. Moreover, almost 17% of female candidates have a woman as their main (i.e., first) letter writer, while this is the case for 11% of male candidates.

We also gather information on the (revealed) quality of the job market paper. More precisely, we consider whether and where the paper has been published, the time to publication and the number of coauthors. These variables are meant to provide a proxy of candidate's quality, which may be visible to letter writers and hiring committees at the time of application. Around 20% of the candidates publish the job market paper in the time period observed, with no difference across genders. No differences across genders appear also in terms of prestige of the publication of the job market paper, nor in terms of number of coauthors or time to publication.

To shed light on subsequent labor market outcomes, we consider both the first placement recorded in the three years after the job market and the current placement. For the subset of candidates for which we have information on the first position, we do not find any gender difference in terms of holding an academic job, with around 80% of male and female candidates in these positions. Also the prestige of the institution of placement and the probability of being an Assistant professor (or higher position) does not display any significant gender difference, whereas female PhDs are more likely to hold a postdoc position.[14] When we turn to analyzing current placement for the full sample of candidates, similarly, we do not find evidence of gender differences in terms of obtaining an academic placement, with about 75% of applicants of both genders in such jobs and, among them, the quality of the job market placement does not display significant gender differences. However, a large gender gap emerges when we consider the position on the academic ladder: male scholars are more than 50% more likely to hold an Associate Professor position, while they are less likely to be Assistant professors or postdocs.

Last, we consider the research output of our pool of candidates. With all the caveats that arise from the literature that we discussed in section 2, these figures suggest that male candidates are more successful during the first years of their academic career in terms of publication records: they have almost one publication more than women, are more likely to publish in one of the Top 5 journals, and their research is more often cited.

We next turn to presenting some descriptive statistics regarding the letter writers in our sample. In particular, Table 2 highlights that the pool of advisors is extremely gender-unbalanced, as only 1,449 out of 8,484 referees are women. The fraction of "ghost" referees who happen to be indicated by a candidate (or some candidates), but never upload their letter(s), is larger among women, potentially suggesting their marginal importance in the students' portfolio. Next, on average female advisors write fewer letters compared to their male counterparts. In line with assortative matching by gender among advisors and students, illustrated above, female sponsors more often tend to work with at least one female PhD student.

Female letter writers appear to lag behind male ones in terms of their research output and career achievement. They are generally more junior, both in terms of career length (the average first year of publication is significantly more recent than that of men) and in terms of academic ranks in that they are less likely to hold a full professor status, consistently also with the leaky pipeline phenomenon in economics. Moreover, they have fewer publications

[14] Other positions include teaching positions within academia and non-academic positions.

in top journals in economics, with nearly half as many citations compared to men.

Table 2: Descriptive statistics of letter writers

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Male letter writer		Female letter writer		Difference		
	mean	sd	mean	sd	Obs	Diff	T-stat
Never uploaded	0.13	0.34	0.16	0.37	8464	-0.029***	(-2.793)
# Letters written	2.66	3.67	1.99	2.83	8464	0.670***	(7.755)
At least 1 female advisee	0.45	0.50	0.52	0.50	7238	-0.065***	(-4.130)
Academic affiliation	0.78	0.42	0.74	0.44	8464	0.034***	(2.700)
Full professor	0.24	0.43	0.19	0.39	8464	0.055***	(4.777)
First publication year	1993.99	12.73	1997.97	10.93	6683	-3.979***	(-10.729)
# articles Repec	19.90	30.45	11.33	17.94	8464	8.567***	(14.395)
# Publications GS	70.80	128.88	46.53	74.58	8464	24.268***	(9.741)
# Top 5 publications	2.24	5.04	1.11	2.51	7860	1.124***	(12.129)
At least 1 Top 5 publication	0.42	0.49	0.34	0.47	7860	0.082***	(5.691)
# Citations	1006.27	2646.35	533.30	1472.93	8464	472.970***	(9.468)
Observations	7015		1449		8464		

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

3.3 Corpus construction and pre-processing

Starting from our sample of 25,778 (potential) references, we exclude those cases in which the referee did not upload the letter even if she was supposed to, and those cases in which the letter provided was in Italian.[15] This leaves us with 21,533 letters.

As our sample includes two different institutions and for the academic one we have data for two departments, there are cases in which the same candidate applied to more than one in the *same* year and turned in the same reference letter. In the analysis, we drop 842 application packages with such duplicate letters, keeping the package with more available letters.

We then anonymize our texts, replacing each reference to the candidate in all letters with the tokens "$candidate\_maleID$" for male and "$candidate\_femaleID$" for female candidates, ID being the individual identifier of each candidate. Moreover, we replace all personal

[15] These are less than 100.

pronouns and determiners (e.g., him, his, her, etc) in the text with such tokens to identify the majority of instances in which the letter refers to the candidate.

We proceed with a standard pre-processing of the text. We first strip off the header and the footer of each letter, since they typically include emails, addresses and affiliation information of the reference letter writer, with no reference to the candidate. Next, we convert the text into lower-case characters, split contractions and remove double spaces, punctuation, numbers and stopwords.[16] We also replace several bi-grams with a single token to simplify the analysis (e.g,. "job market" was replaced by "job-market", "interest rate" by "interest-rate"). All these steps can help a statistical model to only learn from terms that have a relevant meaning, reducing the dimensionality of our corpus.

After having cleaned the text of each letter, we transform it in a list of "tokens", i.e., words or n-grams, and then proceed with its *lemmatization*, i.e., we replace each token with its dictionary base form. All these pre-processing steps allow us to reduce the average length of our documents from 1,029 words in the full letter text to 988 words in the body text (that with no headers or footers) to 536 words after the full pre-processing.

3.4 Corpus description

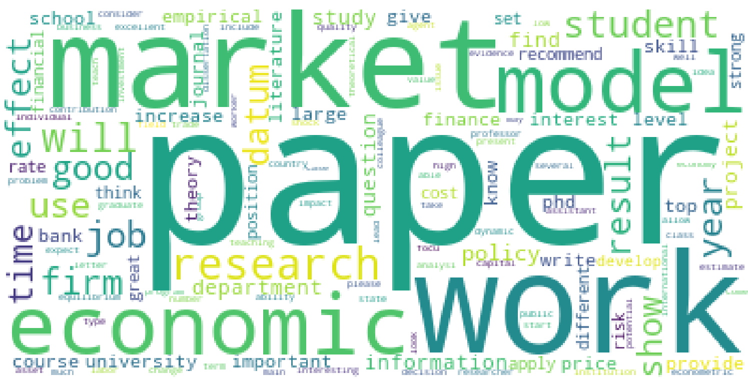
Our final corpus consists of 18,925 documents, which are combinations of 109,744 unique lemmas. The total number of words in our pre-processed documents exceeds 92 millions, while it was over 119 millions before pre-processing.

Our corpus can be represented by a term frequency matrix, which has as many rows as documents ($D = 18,925$) and as many columns as lemmatized words ($V = 109,744$) in our vocabulary. Each element in the matrix will be the frequency in document d of the word v. We can use this information to provide a first graphical representation of our data. Figure 4 displays the most frequent lemmatized words in our corpus, i.e., in all our letters, put together using a word cloud. Unsurprisingly, in our case the most common words are "market" and "paper", as letters mainly discuss the job market paper of the candidate.

In order to extrapolate more meaningful information from the letters, we can weigh the raw frequency of each lemma (its *term frequency, tf*) by the (inverse of the) number of

[16] These are words which do not carry any information per se but rather have some functional purpose, e.g., "the", "to", "of", etc.

Figure 4: Word cloud for the full corpus of reference letters



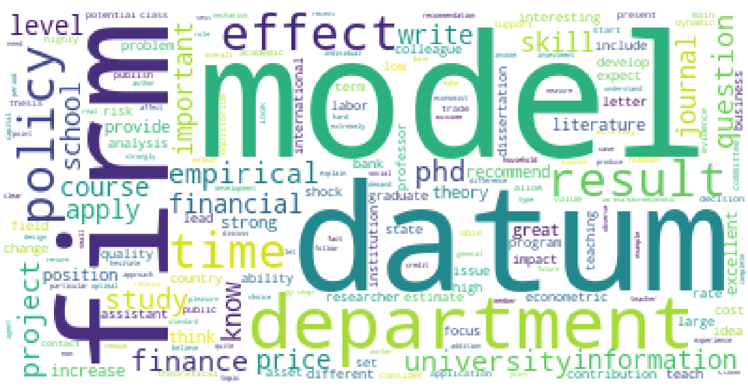
Notes: Word cloud based on raw frequencies of all lemmatized words in the corpus.

documents it occurs in (its *document frequency, df*). This measure is called *tfidf* and, for every lemmatized word v, it is given by:

$$tfidf_v = (1 + log(tf_v)) \times (1 + log\frac{N}{df_v}) \tag{1}$$

Such re-weighting allows us to give low scores to words that occur frequently, but in every document (e.g., function words). Similarly, words that are rare but still appear in most documents in the corpus would also get lower scores. The most prominent example in our setting would be words like "job", "market" and "paper", which were indeed the most frequent ones in Figure 4. Words that are quite frequent but occur only in a few documents get the highest score as these are the words that carry most information about the documents' content. Note also that the use of *logs* dampens the effects of the re-weighting. Figure 5 shows the word cloud that we obtain by reweighting all the words in our corpus by their *tfidf* score. The resulting image is more informative on the content of the letters, highlighting the duality between theory and empirics in the work of the candidates that is described in the letters.

Figure 5: Word cloud with *tfidf* reweighting



Notes: Word cloud based on *tfidf* of words that appear in more than 5% and in less than 75% of documents.

4 Text analysis of reference letters

4.1 Supervised Text Analysis: word embeddings

The description of the corpus provided so far gives little information on what referees say about their students in their letters. Either the use of simple term frequencies or of *tfidf* are not suitable tools to understand how candidates are described. Indeed, these simple descriptive methods do hint at some information regarding the main content of the letters, but this essentially refers to the main topic of the candidate's research.

Our preferred approach to explore how candidates are described in their reference letters will thus be a *supervised* approach. Specifically, we rely on a model with lists of "target words" that likely capture some meaningful characteristics of the candidates. To do so, we build on the literature in psychology described in Section 2. Following Schmader et al. (2007), we start from two categories that have been used to describe job applicants: standout and grindstone terms. They represent, respectively, words referring to the candidates' exceptional character (e.g., outstanding, unique, and exceptional), and words referring to the effort they put in work (e.g., hardworking, conscientious). We then consider two other categories of adjectives that psychologists have identified as often carrying implicit gender stereotypes related to the social role theory. These are agentic and communal adjectives. The first ones refer to personality traits related to self-confidence, assertiveness, tenacity. The latter, instead, refer to personality traits that emphasize a person's ability to sympathize with others (e.g., agreeable, caring, warm). We consider lists (of variable length) for each category according to Schmader et al. (2007), Madera et al. (2009) and Chapman et al. (2020). The full lists are reported in Appendix A.[17]

Having defined such lists of target words, we aim to understand how these are used in reference to candidates. To do so we transform our target words into mathematical objects (i.e., vectors) that represent their semantic meaning using *word embeddings*.[18] This approach identifies words that are most commonly used together, i.e., in a similar context, to capture their relatedness in semantic terms. This idea of semantic relatedness of context, or distributional semantics, is a concept developed in linguistics, dating back to Firth (1957) who stated that "you shall know a word by the company it keeps". Mathematically, this translates into representing each target word as a vector in a low dimensional space, where its position and relative proximity to other words capture their semantic similarity in a way that words with similar meanings or semantically related will lie close together. Note that the dimension of such space will be lower than that of the full vocabulary.

Operationally, our word embedding procedure employs the `word2vec` tool and works as follows. First, we choose the two exogenous parameters to feed into the model. These are the *embedding dimension*, i.e., the dimension of the state space in which we project our text, and the *window size*, i.e., the maximum distance within the text between each word and the target word that defines which tokens are considered. We set the two parameters at 100 and 6, respectively. Moreover, we consider, for computational convenience, only those lemmas that appear at least ten times in our corpus. Second, we estimate our word embedding model, which will produce a vector of 100 dimensions for each of the target words initially identified. The algorithm we use is a *skipgram model*, which computes the probability of observing each context word given the target word we set. The process is iterative: starting from a random embedding, at each iteration the algorithm finds the vectors that minimize a loss function, and then starts again from these vectors. The loss function involves accounting for both the probability of observing each term within the context of the target word and for the probability of *not* observing it. Intuitively, what happens is that at each iteration the word embedding vector becomes more similar to the embeddings of words in its context and

[17] We note that some of the words in the original sources never appear in our corpus, thus they are not reported in the lists, for example, "self-starter", "go-getter", "endearing", etc.

[18] Some recent contributions, namely Caliskan et al. (2017), Kozlowski et al. (2019) and Ash et al. (2023), have used word embeddings in a similar spirit to unveil cultural and gender attitudes.

less similar to the embeddings of words not in its context. After a predetermined number of iterations (100 in our case), the vectors that minimize the loss function will be the optimal word embeddings. These vectors will be our new *vocabulary*.

We measure similarity between the word embeddings of each candidate and of target words by cosine distance, so that word vectors with smaller angles are considered more similar to each other.[19]

In our setting, we compute 42 embeddings, i.e., one for each of the 42 target words. In order to reduce the dimensionality of our vocabulary, we further combine those referring to the words in the same category (standout, grindstone, communal and agentic) to obtain four *average vectors*, one for each category. Once we have transformed our lists of target words into just four mathematical objects given by the average embedding vector of the terms in each category, we compute the cosine distance between these vectors and the (embedded) vectors representing tokens for each candidate (i.e., $candidate\_maleID$ and $candidate\_femaleID$) within a specified corpus. Our first exercise considers all the letters written for a given candidate irrespective of the letter writer. This allows us to obtain a measure of how each candidate is described overall. Table B.3 in the Appendix provides excerpts from letters in which the cosine similarity between references to a candidate and standout/grindstone characterization are high.

4.2 Gender differences in candidate descriptions

We calculate cosine similarity for 6,004 candidates and report them by gender in Table 3. Columns 1 and 3 report the cosine similarity between each personality trait average vector and the target token for, respectively, male and female candidates. Column 5 shows the difference between the two: positive numbers mean that male candidates are more likely to be described in a given way; negative numbers that female candidates are more likely to be described that way. In column 7 we report these differences conditioning on a number of candidates' observable characteristics.

Our findings reveal that advisors tend to describe male students as outstanding more

[19] Cosine similarity can range between -1 and 1. A value of -1 means that the vectors point in diametrically opposite directions, i.e., words have opposite meaning; a value of 1 means that the two vectors point in the same direction, i.e., words are synonyms (or the same word when the vectors exactly overlap); a value of 0 means that the two vectors are orthogonal, i.e., the two words are completely unrelated.

Table 3: Cosine similarity between reference to candidate and target average vectors, by candidate's gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male		Female		Difference			
	mean	sd	mean	sd	Diff	T-stat	Diff (cond)	T-stat
Standout	0.245	0.066	0.240	0.066	0.005***	(2.838)	0.005***	(2.748)
Grindstone	0.216	0.063	0.224	0.064	-0.008***	(-4.538)	-0.005***	(-2.9002)
Communal	0.217	0.064	0.219	0.065	-0.002	(-1.109)	0.002	(1.382)
Agentic	0.236	0.061	0.242	0.061	-0.005***	(-3.066)	-0.001	(-0.637)
Observations	4312		1692		6004		5875	

Notes: The conditional differences in column 7 are computed net of year of application, department to which application was sent, field of research, candidates' PhD institution fixed effect and an indicator for the candidate's job market paper published in a Top 5 journal. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

than they do for female students; also, they tend to stress female candidates' hard-working character more than they do for males. These results are in line with those of Schmader et al. (2007) and persist also controlling for the candidates' year of application, department to which application was sent, field of research, PhD institution fixed effect and an indicator for whether the job market paper was eventually published in a Top 5 journal to proxy for candidate's quality (column 7). Considering then the candidate's assertiveness (agentic traits) vis-à-vis her interpersonal skills (communal traits), the difference in the degree to which both interpersonal skills are stressed is very small and not significant for communal traits. Moreover, the difference in the agentic traits becomes statistically not significant once we condition on the main candidate's observable characteristics.

Taking these pieces of evidence together, we conclude that gender differences are most evident in standout and grindstone categories and, thus, we will focus on them in the regression analysis in Section 5.

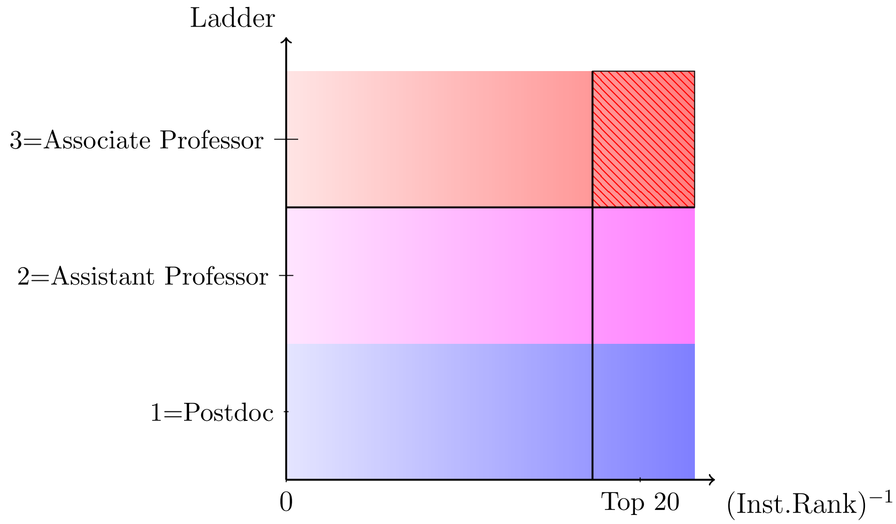
5 Effects on career outcomes

We now turn to examining the labor market outcomes of our sample of job market candidates. In particular, we are interested in understanding how candidates' characterization in letters relate to early and later career achievements, and whether any gender differences appear.

The task of measuring candidates' career success is not straightforward. To start with, there are at least two dimensions of "success", even if we focus on academic placements only: first, the seniority of the position, and, second, the prestige of the institution of affiliation. This is graphically illustrated in Figure 6. The first dimension measures the seniority of the academic titles that job market candidates of different cohorts hold. While one can easily assume that an associate professorship is better than a postdoc position within the same institution, or that an assistant professorship in a Top 5 department is better than an assistant professorship in a lower tier institution, it is hard to compare placements across different seniority levels and institutions. In fact, the second dimension of career success is defined by the prestige of the placement institution. We will thus measure career success not only by the prestige of the placement institution, but also by a "composite" measure that combines the prestige of the institution with its position in the placement ladder.

We start by investigating the first placement of candidates, i.e., the position they obtain within three years from their job market. Our outcome variables of interest are, first, a dummy for being in a Top 20 department according to Repec; second, a dummy for holding an assistant professor position (or higher, as opposed to postdoc or other miscellaneous positions) in a Top 20 department. We then consider their current placement, i.e., career outcomes at the time of our dataset construction, i.e., October 2020-April 2021. These outcomes capture achievement at different stages of the career depending on the candidates' application cohort. Analogous to the first placement, we start from considering as outcome variable a dummy for being in a Top 20 department. We then combine the two dimensions of success depending on the seniority of the candidates: for those candidates who were on the job market more than seven years before the data collection took place, the outcome variable is a dummy for holding an associate professor position in a Top 20 department according to Repec (i.e., we consider whether the candidate's position falls in the upper-right box highlighted in the figure). If, instead, the candidate was on the job market less than seven years before the data collection, the outcome variable is a dummy capturing the probability of having (at least) an assistant professor position. The seven year split accounts for the fact that tenure-track assistant professor positions last on average 6 years. Last, we consider different measures of research output as outcome variables. In particular, we look at the number of publications, whether the candidate has a Top 5 publication, and the number of citations. While first placement is the outcome most directly related to candidates' description in letters, also current outcomes and research output can be indirectly related to them. Although we acknowledge that other factors such as fertility choices may influence current placement and researchers' scientific productivity, we believe it is relevant to explore them as a way to investigate the medium term effects of candidates' characterization in letters.

Figure 6: Ranking over career outcomes



For all outcomes y of candidate i, on the job market in year t, applying to department j,[20] we estimate a pooled linear regression model with the candidate's characterizations in the letters as obtained through word embeddings (WE_{itj}) in Section 4.2 as key explanatory variables. In particular, the cosine similarity between the word embeddings of candidates and standout/ grindstone descriptions are the main variables of interest. We then add three sets of control variables capturing, respectively, candidates' observable characteristics ($Candidate\_X_{itj}$), number of reference letters and main letter writer's observable characteristics ($LetterWriter\_X_{itj}$). Candidates' observable characteristics include gender, the number of publications prior to the job market, the ranking band of the PhD institution (Repec,

[20]Note, indeed, that the same person may appear in our sample more than once if she went on the job market more than once over the period of analysis. Instead, if the same person applied to more than one department in the same year, we keep only one application for that candidate, specifically the one which contains the highest number of reference letters.

seven bands), whether the candidate has published the job market paper in a Top 5 journal, and field of research fixed effects as proxied by the main JEL code the candidate indicated in the application, so as to compare candidates within homogeneous fields of research. Among letter writer's controls we include the number of reference letters and some observable characteristics of the main letter writer including gender, the number of Top 5 publications, and whether she is full professor, to capture seniority and prestige.

The estimation equation is:

$$\begin{aligned} y_{itj} = &\alpha + \beta_1 WE_{itj} + \beta_2 Candidate\_X_{itj} + \\ &\beta_3 LetterWriter\_X_{itj} + \tau_t + \psi_j + \varepsilon_i \end{aligned} \tag{2}$$

In order to account for differences stemming from the candidates' application cohort, we always include in our regressions year of application fixed effects τ_t. This allows to interpret the results as comparisons of candidates of the same "vintage". We also include fixed effects ψ_j for the department of application, to take into consideration potential differences between candidates applying to the three different departments from which we collect data. We start from a parsimonious specification in which we only include among the regressors our key explanatory variable (WE_{itj}), time and department of application fixed effects, to then gradually include candidates' observable characteristics, number of letters and main letter writers' observable characteristics.

5.1 Career success: first placement

Table 4 shows point estimates from a number of simple OLS regressions, where the dependent variable is a dummy for a Top 20 ranked department as the first candidate's placement in columns 1-3 and a composite indicator for the first placement being in a Top 20 ranked department with a position of assistant professor or higher in columns 4-6. In our sample about 11% of the candidates ended up in a Top 20 institution at first placement, 5% in a Top 20 institution as assistant professor or higher.

We start in columns 1 and 4 with a parsimonious specification of equation 2, where we only control for the year and the application department fixed effects, and examine the role of how the candidate is described, by including our measures of similarity to predefined personality characteristics (Section 4.2). In particular, we are interested in the coefficients on the average cosine similarity between each reference to the candidate in the letters and the word embeddings of standout and grindstone words (Schmader et al., 2007). In the interest of brevity, we will often refer to them as letter "standout-ness" or letter "grindstone-ness" in the remainder of the paper. Columns 2 and 5 further controls for gender and for other candidate's characteristics that are meant to capture strength or quality as described above. Columns 3 and 6 finally add controls for the number of letters and the characteristics of the main letter writer.

Results in columns 1 and 4 show that similarity to standout descriptions is positively related to the probability of obtaining a more prestigious placement both in terms of ranking of the department only and of ranking and position combined; on the other hand, similarity to grindstone traits exhibits a negative and statistically significant coefficient. The effects are robust to the inclusion of the different sets of control variables and persist throughout all specifications, with the sole exception of the negative coefficient on grindstone traits in column 6, which becomes marginally not significant. The magnitude of the coefficients, moreover, is non negligible: in our most restrictive specification (columns 3 and 6) a 0.1 increase in the cosine similarity between the word embeddings of standout characterizations and of candidate i is associated with an increase in the likelihood of being hired in a Top 20 institution at first placement of about 2.8 percentage points and with that of being hired in a similar institution as assistant professor or higher of about 2.3 percentage points (relative to the baseline probabilities these increases amount to about, respectively, 26 and 43%). Conversely, a similar 0.1 increase in the cosine similarity to the grindstone embedding is associated with a decrease in our measures of success of about 2 and 1 percentage points (relative to the baseline probabilities, respectively, 18.5 and 17%).

Turning to the coefficients of our sets of control variables, we find that, net of the influence of the letters' content highlighted above: male and female candidates are equally likely to be hired in Top 20 institutions and to be so as assistant professors or higher, once we control for field of specialization fixed effects; the quality of the candidate – as captured by the ranking of the PhD institution attended and the subsequent publication of the job market paper in a Top 5 journal – are positively related with the measures of career success at first placement considered; the same positive relationship emerges with the number of reference letters in the application package and some observable characteristics of the main letter writer such as her number of publications in Top 5 journals and whether she is a full professor (which is statistically significant only for our more demanding measure of success); having a woman as main letter writer has a positive and marginally significant relation with success only for the probability of ending up in a high ranked department, but not when we also consider the position obtained.

Table 4: First placement: probability of being in a Top 20 econ department and of being in a Top 20 department as assistant professor or higher

	(1)	(2)	(3)	(4)	(5)	(6)
	Top 20			Top 20 & Assistant		
Standout cos. sim.	0.367***	0.327***	0.281***	0.293***	0.269***	0.226***
	(0.0703)	(0.0692)	(0.0680)	(0.0549)	(0.0535)	(0.0524)
Grindstone cos. sim.	-0.227***	-0.247***	-0.202***	-0.144**	-0.129**	-0.0908
	(0.0711)	(0.0704)	(0.0706)	(0.0586)	(0.0582)	(0.0582)
Female		0.00894	0.0110		0.0111	0.0124
		(0.00968)	(0.00968)		(0.00785)	(0.00781)
# Publications Pre-JM		-0.000139	-0.000247		0.00181	0.00141
		(0.00215)	(0.00213)		(0.00207)	(0.00203)
Ranking band of PhD inst. (1-7)		-0.0292***	-0.0233***		-0.0156***	-0.0115***
		(0.00246)	(0.00249)		(0.00195)	(0.00198)
Published her job market paper in a top-5		0.236***	0.216***		0.260***	0.247***
		(0.0488)	(0.0483)		(0.0567)	(0.0560)
# Letter writers			0.0309***			0.0269***
			(0.00714)			(0.00603)
Main lett. writer female			0.0247*			-0.00213
			(0.0136)			(0.00931)
# Top 5 public. (main lett. writer)			0.00483***			0.00329***
			(0.000744)			(0.000693)
Full professor (main lett. writer)			0.00815			0.0153**
			(0.00899)			(0.00773)
Mean dependent variable			0.109			0.0531
Year FE	✓	✓	✓	✓	✓	✓
Department FE	✓	✓	✓	✓	✓	✓
JEL code FE		✓	✓		✓	✓
R^2	0.0131	0.0636	0.0841	0.0141	0.0627	0.0848
N	5095	5093	5093	3956	3956	3956

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

5.2 Career success: current placement

Table 5 shows results for specifications analogous to those of Table 4, using current placement outcomes, in terms of probability of holding (any) position at a Top 20 institution at the time we collected the information (columns 1-3), and of holding an associate or higher position (an assistant professor or higher position) at a Top 20 institution for candidates who applied more (less) than seven years before the scraping of the career data (columns 4-6).

Our findings show that the characterization of a candidate in reference letters still significantly correlates with career outcomes several years after the job market. In both tables, indeed, we observe that the standout descriptions have a positive effect on current career success, while the opposite holds for grindstone characterizations, whose use in describing candidates penalizes their career success. As expected, the magnitude of the coefficients for how candidates are talked about is – though just slightly – smaller than that estimated for first placement. Indeed, we imagine that reference letters weigh more at the onset of the career. In the most restrictive specifications (columns 3 and 6), a 0.1 increase in standout cosine similarity raises the probability of currently working in a Top 20 institution and that of doing so with a high rank position by, respectively, 2 and 1.75 percentage points, i.e., 23% and 34% over the baseline. As for the grindstone coefficient, a similar 0.1 increase corresponds to a decrease in the likelihood of professional success as measured through our two proxies, in the order of 17.5 and 8.5 percentage points, respectively 21 and 17% over the baseline. Other observable characteristics included in the regression analysis remain roughly similar in terms of sign and statistical significance to those discussed in the previous section, with an exception for the (conditional) gender gap, which reveals that women are significantly more likely to be working in a highly ranked institution when we do not consider their position.

Overall, we have presented evidence on a strong and robust relationship between “standoutness” and “grindstone-ness” in candidates’ characterizations in reference letters and career outcomes, either observed at the onset of PhD students’ academic careers or a few years later. These patterns hold accounting for a number of observable candidate, advisor and application package characteristics, and taking into account year, department of application and field codes fixed effects.

Table 5: Current placement: probability of being in a Top 20 econ department and of being in a Top 20 department as associate or assistant professor or higher (vintage based)

	(1)	(2)	(3)	(4)	(5)	(6)
	Top 20			Top 20 & Assist. or Assoc.		
Standout cos. sim.	0.261***	0.237***	0.199***	0.223***	0.207***	0.175***
	(0.0582)	(0.0581)	(0.0572)	(0.0455)	(0.0453)	(0.0441)
Grindstone cos. sim.	-0.203***	-0.217***	-0.175***	-0.121***	-0.122***	-0.0855**
	(0.0575)	(0.0580)	(0.0572)	(0.0430)	(0.0436)	(0.0429)
Female		0.0163**	0.0187**		0.00188	0.00387
		(0.00822)	(0.00817)		(0.00627)	(0.00621)
# Publications Pre-JM		-0.000128	-0.000238		-0.000726	-0.000818
		(0.00181)	(0.00181)		(0.00126)	(0.00126)
Ranking band of PhD inst. (1-7)		-0.0230***	-0.0175***		-0.0155***	-0.0110***
		(0.00209)	(0.00206)		(0.00167)	(0.00163)
Published her job market paper in a top-5		0.213***	0.190***		0.209***	0.190***
		(0.0468)	(0.0462)		(0.0446)	(0.0441)
# Letter writers			0.0314***			0.0285***
			(0.00606)			(0.00498)
Main lett. writer female			0.0120			0.00735
			(0.0109)			(0.00838)
# Top 5 public. (main lett. writer)			0.00441***			0.00358***
			(0.000644)			(0.000576)
Full professor (main lett. writer)			0.0173**			0.0121*
			(0.00765)			(0.00625)
Mean dependent variable			0.0821			0.0505
Year FE	✓	✓	✓	✓	✓	✓
Department FE	✓	✓	✓	✓	✓	✓
JEL code FE		✓	✓		✓	✓
R^2	0.0124	0.0541	0.0785	0.0122	0.0512	0.0774
N	5700	5698	5698	5700	5698	5698

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

5.3 Research output

In this section we explore whether candidate characterization in reference letters in terms of standout and grindstone also reflects onto early research outcomes.

In Table 6 we thus estimate our model in equation (2) on different measures of research output, namely, the (log 1+) number of publications (columns 1 and 2), an indicator variable for whether the candidate has any publication in one of the Top 5 journals (columns 3 and 4), and the (log 1+) number of citations (columns 5 and 6). We exclude from the regressors capturing the candidate's quality the number of publications prior to the job market and the indicator for the job market paper being published in a Top 5 journal.

Table 6: Other outcomes: publication records

	(1)	(2)	(3)	(4)	(5)	(6)
	# publications		Top5 publication		# citations	
Standout cos. sim.	0.480***	0.416**	0.173***	0.142***	1.357***	1.229***
	(0.170)	(0.170)	(0.0503)	(0.0497)	(0.382)	(0.380)
Grindstone cos. sim.	-0.290*	-0.144	-0.123**	-0.120**	-1.286***	-0.967**
	(0.172)	(0.175)	(0.0505)	(0.0509)	(0.396)	(0.398)
Female		-0.169***		-0.0181***		-0.271***
		(0.0222)		(0.00646)		(0.0519)
Ranking band of PhD inst. (1-7)		0.0147**		-0.00516***		-0.000550
		(0.00578)		(0.00179)		(0.0133)
# Letter writers		0.0692***		0.0126**		0.208***
		(0.0183)		(0.00507)		(0.0404)
Main lett. writer female		0.0221		0.00137		0.0926
		(0.0319)		(0.00896)		(0.0719)
# Top 5 public. (main lett. writer)		0.000940		0.00305***		0.0141***
		(0.00138)		(0.000590)		(0.00350)
Full professor (main lett. writer)		0.0331		0.00648		0.0656
		(0.0219)		(0.00659)		(0.0497)
Mean dependent variable		0.628		0.0637		1.449
Year FE	✓	✓	✓	✓	✓	✓
Department FE	✓	✓	✓	✓	✓	✓
JEL code FE		✓		✓		✓
R^2	0.113	0.134	0.0353	0.0662	0.119	0.152
N	6000	6000	6000	6000	6000	6000

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%. The number of publications and the number of citations are in logarithms $(\# + 1)$.

The signs of the variables measuring the letter emphasis on candidates' standout and grindstone characteristics continue to consistently indicate the same pattern, with the former (the latter) being positively (negatively) associated with different measures of research productivity. Only for the first outcome considered, our estimated coefficient for grindstonness is not statistically significant.

Naturally, these relationships may be partially confounded by other factors along the candidate's career, such as collaboration patterns, research interests, or promotion decisions. However, we see this evidence as suggesting that the disadvantage to women that is generated by the way they are described in their reference letters on the junior job market reflects – arguably through initial placement – also onto longer term career outcomes.

5.4 Robustness checks

In this section we provide several robustness checks to corroborate our main estimates. First, in a further attempt to isolate more neatly the quality of the candidate, we replicate the regressions in columns 3 and 6 of Tables 4 and 5 including PhD institution fixed effects (rather than PhD institutions' ranking bands) in the regression. This accounts for unobservable differences in candidates' quality, since candidates admitted and graduating from the same institution and within the same cohort are likely subject to similar selection and promotion standards, making them of more similar quality. The results are shown in Table 7 and are in line with those of Tables 4 and 5.

In Table 8 we then investigate the robustness of our findings to alternative definitions of early (panel A) and current (panel B) career outcomes. Namely, we restrict or expand the group of prestigious institutions, while keeping fixed our measures of academic ladder. In particular, we consider placement at a Top 10 and Top 50 institution in columns 1 and 2 and the composite measure of early (current) career success that considers a higher ranked position at a Top 10 and Top 50 institution in columns 3 and 4. The regression specification corresponds to the most stringent one, as those used in columns 3 and 6 of Table 4 and 5. The results confirm that a higher similarity to standout characterizations is positively related to all measures of career success we consider, the broader the definition of success, the larger the estimated coefficients. For grindstone, indeed, the estimated coefficients become not significant when we look at current placement in very highly ranked departments.

Table 7: Robustness: Career success at first and current placement with PhD institution FE

	(1)	(2)	(3)	(4)
	First placement		Current placement	
	Top 20	Top 20 Assist.	Top 20	Top 20 & Assist. or Assoc.
Standout cos. sim.	0.249***	0.203***	0.138**	0.142***
	(0.0743)	(0.0593)	(0.0623)	(0.0491)
Grindstone cos. sim.	-0.208**	-0.112*	-0.183***	-0.115**
	(0.0811)	(0.0675)	(0.0659)	(0.0522)
Mean dependent variable	0.111	0.0540	0.0827	0.0513
Year FE	✓	✓	✓	✓
Department FE	✓	✓	✓	✓
JEL code FE	✓	✓	✓	✓
PhD Institution FE	✓	✓	✓	✓
Candidate X_{ijt}	✓	✓	✓	✓
Letter Writer X_{ijt}	✓	✓	✓	✓
R^2	0.173	0.148	0.151	0.137
N	4979	3850	5573	5573

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

6 Mechanisms and discussion

Our analysis has unveiled that there is a gender gap in how candidates are described in their reference letters, with letters emphasizing men's being brilliant and women's being diligent and hardworking. Moreover, we showed that these descriptions of candidates influence their subsequent career outcomes both in the short run (at first placement) and in the longer run (looking at their current placement and at their publication records).

The question we now turn to regards the drivers behind gendered patterns in candidates' descriptions, since they ultimately may be relevant for policy implications. In this section we thus attempt to discuss the potential underlying mechanisms and present some evidence on their role behind the gendered use of standout and grindstone descriptions, and the positive (negative) value standout (grindstone) characterizations bring to Economics PhD candidates' career success.

First, the observed gendered patterns in candidates' descriptions may be driven by the supply side, i.e., it may be that male and female candidates are intrinsically different in

Table 8: Robustness: alternative definitions of career outcomes

	(1)	(2)	(3)	(4)
A. First placement	Top 10	Top 50	Top 10 & Assist.	Top 50 & Assist.
Standout cos. sim.	0.186***	0.454***	0.127***	0.431***
	(0.0567)	(0.0943)	(0.0394)	(0.0815)
Grindstone cos. sim.	-0.178***	-0.294***	-0.0801*	-0.149*
	(0.0582)	(0.0952)	(0.0423)	(0.0859)
Mean dependent variable	0.0713	0.220	0.0298	0.124
B. Current placement	Top 10	Top 50	Top 10 & Assist. or Assoc.	Top 50 & Assist. or Assoc.
Standout cos. sim.	0.0806*	0.425***	0.0855***	0.374***
	(0.0466)	(0.0817)	(0.0331)	(0.0672)
Grindstone cos. sim.	-0.0713	-0.324***	-0.0301	-0.227***
	(0.0450)	(0.0825)	(0.0304)	(0.0698)
Mean dependent variable	0.0479	0.181	0.0260	0.123
Year FE	✓	✓	✓	✓
Department FE	✓	✓	✓	✓
JEL code FE	✓	✓	✓	✓
Candidate X_{ijt}	✓	✓	✓	✓
Letter Writer X_{ijt}	✓	✓	✓	✓
R^2	0.0404	0.0630	0.0447	0.0746
N	5699	5699	5699	5699

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

terms of their being brilliant or hard-working. We argue, however, that if this was the case, we would expect both male and female referees to characterize them differently.

To investigate potential differences in how letter writers of different genders describe male and female job market candidates, we proceed as follows. First, we classify texts according to who wrote the letter and to the gender of the person the letter was written for. This partitions our corpus in documents, each of which is formed of the set of letters written by a given referee for all his/her male advisees or the set of letters he/she wrote for all his/her female advisees. Relative to our main analysis (Section 4.1), the number of documents we analyze with this method is thus larger, but each document will be shorter than in our main analysis.[21]

Operationally, we substitute each reference to candidate with a token identifying the referee identity and the candidate's gender ($candidate\_male\_refID$, $candidate\_female\_refID$) and then compute the cosine similarity between the average vectors for the target words, i.e., standout and grindstone, and the embeddings corresponding to these new identifiers. Hence, for each referee, we obtain separate measures of cosine similarity to each target semantic category for male candidates and for female candidates.[22] Our approach thus differs from the candidate-level analysis in Section 4.1, because in that case we measured the semantic similarity between the target words and the reference to a single candidate irrespective of who wrote the letter, whereas in this case, between the target words and references to candidates of different genders for whom each referee has written letters. After this exercise, we obtain at most two cosine similarity measures for each referee.[23]

Table 9 compares these measures across candidates' and letter writers' gender in a simple regression framework, mimicking that of a simple Difference-in-differences approach. Namely, in order to investigate if differences in characterizations of male and female candidates are different among male and female sponsors, we regress standout (grindstone) cosine similarities on candidate and letter writer gender and on the interaction between the two terms. Columns 1 and 4 show the results from a parsimonious specification with no control variables, while columns 2 and 5 include letter writer control variables and columns 3 and 6

[21] This additional analysis comes at the cost of potentially reducing the precision of our word embeddings.

[22] Only one measure can be computed for those letter writers who wrote references for candidates of only one gender.

[23] In a separate exercise, we classify text according to the gender of the letter writer, without making a distinction by candidate gender. Operationally, these consists of substituting each reference to candidate with a token identifying the referee identity ($candidate\_refID$), regardless of the candidate's identity. This provides, for example, the measures of the "average" (across candidates) referee "standoutness" or "grindstoneness", i.e., whether each referee is more likely to talk about both male and female candidates using standout or grindstone words in the letters he/she writes. Table B.2 in the Appendix reports the cosine similarity between each personality trait average vector and references to candidates (of both gender), distinguishing between male and female letter writers, and tests for the presence of gender differences along this dimension. It shows that female letter writers tend to emphasize (all) candidate personality traits more, a difference that persists also accounting for several observable referee characteristics (i.e., affiliation institution fixed effects, indicators for having an academic affiliation, being full professor, having at least one female advisee). All in all, this indicates that female advisors may provide more information on personal characteristics of the candidates, beside their professional achievements.

Table 9: Candidate characterizations by candidate and letter writer gender

	(1)	(2)	(3)	(4)	(5)	(6)
	Standout cos. sim.			Grindstone cos. sim		
Female candidate	-0.00414**	-0.00373**	-0.00265	0.00647***	0.00843***	0.0131***
	(0.00180)	(0.00178)	(0.00199)	(0.00177)	(0.00174)	(0.00189)
Female letter writer	-0.00706**	-0.00556**		0.0159***	0.0141***	
	(0.00280)	(0.00277)		(0.00275)	(0.00270)	
Female candidate × Female letter writer	0.00621	0.00474	0.00691	-0.00136	-0.00376	-0.0129***
	(0.00440)	(0.00431)	(0.00525)	(0.00432)	(0.00420)	(0.00477)
Letter Writer X_{ijt}		✓			✓	
Letter Writer FE			✓			✓
R^2	0.00124	0.126	0.656	0.00830	0.144	0.699
N	8405	8405	3950	8405	8405	3950

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

letter writer fixed effects.[24]

The results in the first line confirm that female candidates are systematically described as less standout and more grindstone. However, and most importantly, the estimated differences in candidates' characterizations across genders are more pronounced in the case of male letter writers: the coefficient of interest on the interaction term is always of the opposite sign compared to that on *Female candidate*. For those advisors who work with students of both genders, we calculate these differences holding constant the letter writer identity (column 3 and 6). The patterns we detect are confirmed and highly significant in the case of grindstone cosine similarity: a given male letter writer appears to describe his students of different gender differently, putting more focus on grindstone characteristics when referring to female students, while female letter writers are virtually gender neutral in using such characterization. Indeed, the magnitude of the interaction coefficients is in most specifications such as to cancel out the effect detected in the first line. While not statistically significant in the case of standout, our estimated coefficients remain remarkably stable across specifications.

Overall, this evidence indicates that male referees are more prone to the use of gendered language relative to female referees (e.g., only male advisors describe female candidates as more grindstone, whereas female advisors do not) and holds even when we control for the

[24]Note that this specification is very stringent and relies only on a smaller sample of referees who wrote letters for candidates of both genders.

letter writer fixed effects, which accounts for any possible selection of students across referees of different gender. We believe that this evidence runs against the hypothesis that supply side factors, i.e. intrinsic gender differences among candidates, explain the more frequent male candidate description as brilliant and female candidate description as diligent and hard-working.

Alternatively, candidate descriptions in terms of grindstone or standout adjectives may differ by gender because letter writers anticipate that these characteristics will be demanded on the job market and rewarded differently according to gender. For example, if gender norms prescribe women to be more dutiful and hardworking, as highlighted in Babcock et al. (2017), PhD advisors may load these characteristics in their letters hoping to favor their placement. To discuss this hypothesis, in Table 10 we investigate the presence of gender differences in the relationship between standout and grindstone characterizations and career success of candidates, measured both at first (panel A) and at current placement (panel B). In Columns 1 and 3 we replicate the most parsimonious specification, e.g., as in columns 1 and 4 of Tables 4 and 5, and add an interaction between an indicator for female candidates and the two key variables, i.e., the similarity to standout and grindstone descriptions. In columns 2 and 4, we estimate the model with the full set of controls, as in columns 3 and 6 of Tables 4 and 5, and the gender interaction between the candidate characterization variables of our interest.

The results suggest that standout characterizations are beneficial only to male candidates: indeed, the interaction term with the indicator *Female* is negative and virtually almost cancels out the main positive effect detected for men. The coefficients associated with the interaction between the female dummy and candidates' grindstone cosine similarity, though not statistically significant, are always negative, suggesting that the effect of being described as hard-working is, at a minimum, equally detrimental to men and women. The results are robust across all specifications and all definitions of career success. This allows us to rule out that women are described as "grindstone" and not "standout" to comply to gender norms, because they do not benefit from such characterization.

All in all, the evidence presented above is inconsistent with differences in how male and female candidates are described being driven by either intrinsic differences across candidates (supply side), or by compliance to gender norms (demand side). We argue that these differences may instead capture implicit gender biases whereby senior male academics and

Table 10: Career outcomes, effects by candidate gender

	(1)	(2)	(3)	(4)
A. First placement	Top 20		Top 20 & Assist.	
Standout cos. sim.	0.465***	0.360***	0.373***	0.300***
	(0.0862)	(0.0826)	(0.0659)	(0.0638)
Grindstone cos. sim.	-0.210**	-0.171**	-0.140**	-0.0798
	(0.0847)	(0.0835)	(0.0699)	(0.0680)
Female	0.101**	0.0985**	0.0723**	0.0774**
	(0.0415)	(0.0407)	(0.0324)	(0.0318)
Female × Standout cos. sim.	-0.331**	-0.272*	-0.267**	-0.255**
	(0.146)	(0.143)	(0.117)	(0.113)
Female × Grindstone cos. sim.	-0.0606	-0.0993	-0.00615	-0.0181
	(0.155)	(0.150)	(0.127)	(0.123)
Mean dependent variable		0.109		0.0533
B. Current placement	Top 20		Top 20 & Assist. or Assoc.	
Standout cos. sim.	0.356***	0.273***	0.298***	0.238***
	(0.0703)	(0.0685)	(0.0564)	(0.0547)
Grindstone cos. sim.	-0.175***	-0.136**	-0.0899*	-0.0493
	(0.0668)	(0.0657)	(0.0524)	(0.0514)
Female	0.113***	0.108***	0.0831***	0.0819***
	(0.0353)	(0.0346)	(0.0247)	(0.0245)
Female × Standout cos. sim.	-0.306**	-0.257**	-0.256***	-0.217**
	(0.123)	(0.122)	(0.0890)	(0.0878)
Female × Grindstone cos. sim.	-0.112	-0.124	-0.0985	-0.115
	(0.132)	(0.129)	(0.0912)	(0.0894)
Mean dependent variable		0.0821		0.0505
Year FE	✓	✓	✓	✓
Department FE	✓	✓	✓	✓
JEL code FE		✓		✓
Candidate X_{ijt}		✓		✓
Letter Writer X_{ijt}		✓		✓
R^2	0.0144	0.0921	0.0155	0.0963
N	5094	5093	3942	3942

Notes: * denotes significance at 10%, ** significance at 5% and *** significance at 1%.

professionals hold an implicit belief that (good) female candidates are hardworking scholars, whereas (good) male candidates are brilliant ones. Even more, these differences are not driven by advisor selection, as they are present also "within the same advisor" (Table 9, column 6).

7 Conclusions

The goal of this paper is to estimate whether and to what extent female graduate students are subject to different reference letter writing practices and how the latter relate to success of their career. In particular, we analyze gender differences in how candidates are described by their sponsors and assess the presence of implicit gender stereotypes conveyed by the language used.

To these ends, we built a novel dataset containing information on job market candidates applying to two top institutions hiring on the international market for junior economist positions. Our analysis combines information on demographic characteristics and labor market outcomes with an innovative set of measures built through the text analysis of the candidates' reference letters.

Our findings reveal significant gender differences in the way male and female job market candidates are presented on the job market. Differences concern not only some observable factors, such as the likelihood of having a female advisor, or the number of sponsors, but also how they are characterized in reference letters by senior academics. In particular, we find that female candidates are consistently described more in terms of being diligent and hardworking rather than outstanding or brilliant. Linking this information with proxies of career success, we find that such features relate to candidates' placement by lowering the success of female PhD graduates in early and current career outcomes and publication records.

Interestingly, we show that differences in candidates' descriptions are driven by letters written by male sponsors, whereas women make no differences based on the gender of the candidate. Moreover, female and male candidates have different returns to the way they are described: in particular, women get almost no benefit from being described in stand-out terms, whereas they are harmed when described in grindstone terms. We argue that this evidence alarms about the potential presence of implicit gender stereotypes behind the observed characterizations. Policy-wise, this highlights a potential structural flaw in the academic job market process that, by heavily relying on reference letters, effectively puts female candidates in a weaker position to compete.

More broadly, our research contributes to a better assessment of the use of referral processes in the labor market. In fact, the use of references is by no means limited to academia. For instance, performance reviews are key tools in organizations to evaluate an employee performance and, while they have the advantage of setting goals and design career trajectories, their language could be influenced by subjective impressions of managers/evaluators and reflect implicit stereotypes on the appropriate characteristics and roles of men and women. We illustrate, indeed, that labor market appraisals do contain gendered language with potential consequences for the career paths of male and female professionals. These patterns may be especially relevant in contexts that are highly male dominated, as suggested by the finding that gender stereotypes are particularly present among male advisors.

Raising awareness on these issues may help on two levels. At the personal level, letter writers' attitudes and behaviors may change when their own biases are revealed to them (e.g., Carlana, 2019 and Boring and Philippe, 2021). At the institutional level, it may help restructuring the referral process to make it less prone to gender stereotypes, for instance by limiting it to the use of closed-form questions.

A Word lists

To obtain the average vectors that characterize each of the semantic categories described in Section 4.2, we adopt the lists used in the literature (we start from Schmader et al. (2007) for the first two categories and from Chapman et al. (2020) for the last two). Below we report the full lists of words in each category.

- **Standout Adjectives:** ["standout", "best", "leader", "exceptional", "outstanding", "star", "superstar", "impressive"]
- **Grindstone Adjectives:** ["hardworking", "tenacious", "deliberate", "productive", "efficient"]
- **Communal Adjectives:** ["likable", "friendly", "enthusiastic", "enthusiasm", "agreeable", "caring", "nice", "pleasant", "kind", "kindness", "warm", "warmth", "cheerful", "polite", "smile", "modest", "humble", "genuine", "collaborative", "upbeat"]
- **Agentic Adjectives:** ["able", "competitive", "proactive", "accomplished", "energetic", "eager", "ambitious", "ambition", "confident"]

B Additional figures and tables

Table B.1: Descriptive statistics of job market candidates on the European Job Market, 2020/2021.

	N	Male	Female	Difference
American/Canadian PhD	787	0.438	0.416	0.022
EU PhD	787	0.436	0.490	-0.054
Italian PhD	787	0.033	0.049	-0.016
Applied micro	787	0.515	0.671	-0.156***
Macro/International/Finance	787	0.210	0.156	0.053*
Theory/Quantitative	787	0.193	0.136	0.057*
Phd Uni Top20 (QS)	787	0.149	0.132	0.017
Phd Uni Top20 Econ	787	0.256	0.198	0.058*
Observations	787			

Notes: Elaborations on data from the European Economic Association job market candidates directory. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Table B.2: Cosine similarity between reference to candidate and target average vectors, by referee's gender

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Male referee		Female referee		Difference (uncond.)			Difference (cond.)		
	mean	sd	mean	sd	Obs	Diff	T-stat	Obs	Diff	T-stat
Standout	0.237	0.072	0.237	0.074	7097	0.001	(0.390)	6845	0.00006	(0.02)
Grindstone	0.195	0.069	0.210	0.070	7097	-0.016***	(-7.121)	6845	-0.014***	(-6.43)
Communal	0.189	0.074	0.195	0.077	7097	-0.006***	(-2.612)	6845	-0.005**	(-2.15)
Agentic	0.213	0.063	0.225	0.062	7097	-0.012***	(-5.856)	6845	-0.011***	(-5.64)
Observations	5916		1181		7097			6845		

Notes: * $p < 0.1$, ** $p < .05$, *** $p < 0.01$. The conditional difference in column 9 accounts for indicators for those with an academic affiliation, with full professorship and with at least one female advisee and for the letter writer affiliation institution fixed effects.

Table B.3: Letter Samples

High grindstone, low standout candidate	
Lines 7-8	Candidate is a bright, motivated and a hard-working researcher. Candidate is very careful, meticulous and thorough with Candidate work.
Lines 14-16	[...] one of Candidate great strengths is an uncommon ability to discover interesting datasets (an obvious basic ingredient for a successful empirical economist) and preview its potential in delivering academically interesting results
Line 24	Candidate has a range of research area and projects.
Lines 39-44	Candidate is a very productive researcher with a lot of passion and motivation: Candidate has energy and an extraordinary intellectual curiosity for relevant economic questions with a twist of policy interest. Candidate portfolio of papers is likely to evolve into a solid research agenda in economics. Candidate is also a nice person, easy to interact with, cooperative and gentle. Candidate would be a great colleague to have around.
Low grindstone, high standout candidate	
Lines 1-2	Candidate is one of the top rated XX economists coming out of XX, and Candidate is very likely to be a star on this year's job market.
Lines 4-9	Candidate impressed me in class [...] this brings back fond memories of my conversations with Candidate about that paper and Candidate research in general [...] The point here is Candidate ability to impress with fundamental insights. This is Candidate hallmark, one that only grows over time.
Line 18	It is here that Candidate makes some fundamental contributions.
Line 61	Candidate is quite prolific and has an exciting research agenda.
Lines 65-68	Candidate's work is filled with insights. Candidate is a deep thinker and will be a wonderful, interactive colleague. Candidate has broad interests and an exciting agenda. I will miss Candidate collaboration as Candidate has become a full-fledged colleague. I am recommending Candidate to all schools including those at the very top.

Notes: The table reports excerpts from reference letters with sentences used to characterize two candidates, with high(low) grindstone and low(high) standout characterizations.

References

Ash, E., D. L. Chen, and A. Ornaghi (2023). Gender Attitudes in the Judiciary:Evidence from U.S. Circuit Courts. *American Economic Journal: Applied Economics* (forthcoming).

Auriol, E., G. Friebel, A. Weinberger, and S. Wilhelm (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 119*(16), e2118853119.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review 107*(3), 714–747.

Bayer, A. and C. E. Rouse (2016). Diversity in the economics profession: A new attack on an old problem. *Journal of Economic Perspectives 30*(4), 221–42.

Bello, P., A. Casarico, and D. Nozza (2023). Research Similarity and Women in Academia. Working paper 10657, CESifo.

Belot, M., M. Kurmangaliyeva, and J. Reuter (2023). Gender Diversity and Diversity of Ideas. Discussion paper 51784-1701087467, CEPR.

Bleemer, Z. (2019). Gender stereotypes in professor-student interactions.

Boring, A. and A. Philippe (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics 193*, 104323.

Bostwick, V. K. and B. A. Weinberg (2020). Peer effects in graduate programmes. In S. Lundberg (Ed.), *Women in Economics*, Chapter 8, pp. 65–71. London: CEPR Press.

Boustan, L., A. Langan, and I. B. Palmer (2020). Variation in women's success across phd programmes in economics. In S. Lundberg (Ed.), *Women in Economics*, Chapter 7, pp. 57–64. London: CEPR Press.

Caliskan, A., J. J. Bryson, and A. Narayanan (2017). Semantics derived automatically from language corpora contain human-like biases. *Science 356*(6334), 183–186.

Card, D., S. D. Vigna, P. Funk, and N. Iriberri (2020). Are Referees and Editors in Economics Gender Neutral? *The Quarterly Journal of Economics 135*(1), 269–327.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *The Quarterly Journal of Economics 134*(3), 1163–1224.

Chapman, B., M. Rooney, E. Ludmir, D. Cruz, A. Salcedo, C. Pinnix, P. Das, R. Jagsi, C. Thomas, and E. Holliday (2020). Linguistic biases in letters of recommendation for radiation oncology residency applicants from 2015 to 2019. *Journal of Cancer Education*.

Chari, A. (2023, May). Report: Committee on the status of women in the economics profession (cswep). *AEA Papers and Proceedings 113*, 815–39.

Dupas, P., A. S. Modestino, M. Niederle, J. Wolfers, and T. S. D. Collective (2021). Gender and the Dynamics of Economics Seminars. NBER Working Papers 28494, National Bureau of Economic Research, Inc.

Dutt, K., D. Pfaff, A. Bernstein, J. Dillard, and C. Block (2016). Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience 9*.

Eagly, A., W. Wood, and A. Diekman (2000). *Social role theory of sex differences and similarities: A current appraisal*, pp. 123–174. Erlbaum.

Eagly, A. H. and M. C. Johannesen-Schmidt (2001). The leadership styles of women and men. *Journal of Social Issues 57*(4), 781–797.

Eberhardt, M., G. Facchini, and V. Rueda (2023). Gender Differences in Reference Letters: Evidence from the Economics Job Market. *The Economic Journal 133*(655), 2676–2708.

European Commission (2021). SHE Figures 2021. Technical report, Directorate-General for Research and Innovation.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-55. *Studies in Linguistic Analysis (special volume of the Philological Society) 1952-59*, 1–32.

Fortin, N., T. Lemieux, and M. Rehavi (2021). Gender differences in fields of specialization and placement outcomes among phds in economics. *AEA Papers and Proceedings 111*, 74–79.

Gaulé, P. and M. Piacentini (2018). An advisor like me? advisor gender and post-graduate careers in science. *Research Policy 47*(4), 805–813.

Haney-Lopez, I. (2000). Institutional racism: Judicial conduct and a new theory of racial discrimination. *The Yale Law Journal 109*, 1717.

Hengel, E. (2022, 05). Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review. *The Economic Journal 132*(648), 2951–2991.

Hilmer, C. and M. Hilmer (2007). Women helping women, men helping women? same-gender mentoring, initial job placements, and early career publishing success for economics phds. *American Economic Review 97*(2), 422–426.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology 65*(1), 373–398.

Jansson, J. and B. Tyrefors (2022). Grading bias and the leaky pipeline in economics: Evidence from stockholm university. *Labour Economics 78*, 102212.

Janys, L. (2022). Evidence for a Two-Women Quota in University Departments across Disciplines. *Review of Economics and Statistics Forthcoming*.

Kim, S. D. and P. Moser (2021). Women in science. lessons from the baby boom. Working Paper 29436, National Bureau of Economic Research.

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

Koffi, M. (2021b). Innovative ideas and gender inequality. CLEF Working Paper Series 35, Canadian Labour Economics Forum (CLEF), University of Waterloo.

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review 84*(5), 905–949.

Lundberg, S. (2020). *Women in Economics.* A VoxEU.org ebook. CEPR Press.

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

Madera, J. M., M. R. Hebl, and R. C. Martin (2009). Gender and letters of recommendation for academia: agentic and communal differences. *Journal of Applied Psychology 94(6)*, 1591–9.

Neumark, D. and R. Gardecki (1998). Women helping women? role model and mentoring effects on female ph.d. students in economics. *Journal of Human Resources 33*(1), 220–246.

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

Paredes, V., M. D. Paserman, and F. J. Pino (2023). Does Economics Make You Sexist? *The Review of Economics and Statistics*, 1–47.

Pezzoni, M., J. Mairesse, P. Stephan, and J. Lane (2016). Gender and the publication output of graduate students: A case study. *PloS one 11*.

Rizzica, L. (2013). Home or away? Gender differences in the effects of an expansion of tertiary education supply. Questioni di Economia e Finanza (Occasional Papers) 181, Bank of Italy, Economic Research and International Relations Area.

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

Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2021). Gender differences in recognition for group work. *Journal of Political Economy 129*(1), 101–147.

Sarsons, H. and G. Xu (2021). Confidence men? evidence on confidence and gender among top economists. *AEA Papers and Proceedings 111*, 65–68.

Schmader, T., J. Whitehead, and V. Wysocki (2007). A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex roles 57*, 509–514.

Trix, F. and C. Psenka (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse & Society 14*(2), 191–220.

RECENTLY PUBLISHED "TEMI" (*)

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2022

ANDINI M., M. BOLDRINI, E. CIANI, G. DE BLASIO, A. D'IGNAZIO and A. PALADINI, *Machine learning in the service of policy targeting: the case of public credit guarantees,* Journal of Economic Behavior & Organization, v. 198, pp. 434-475, **WP 1206 (February 2019).**

ANGELICO C., J. MARCUCCI, M. MICCOLI and F. QUARTA, *Can we measure inflation expectations using twitter?,* Journal of Econometrics, v. 228, 2, pp. 259-277, **WP 1318 (February 2021).**

BARTOCCI A., A. NOTARPIETRO and M. PISANI, *Covid-19 shock and fiscal-monetary policy mix in a monetary union,* Economic challenges for Europe after the pandemic, Springer Proceedings in Business and Economics, Berlin-Heidelberg, Springer, **WP 1313 (December 2020).**

BOTTERO M., C. MINOIU, J. PEYDRÒ, A. POLO, A. PRESBITERO and E. SETTE, *Expansionary yet different: credit supply and real effects of negative interest rate policy,* Journal of Financial Economics, v. 146, 2, pp. 754-778, **WP 1269 (March 2020).**

BRONZINI R., A. D'IGNAZIO and D. REVELLI, *Financial structure and bank relationships of Italian multinational firms,* Journal of Multinational Financial Management, v. 66, Article 100762, **WP 1326 (March 2021).**

CANTELMO A., *Rare disasters, the natural interest rate and monetary policy,* Oxford Bulletin of Economics and Statistics, v. 84, 3, pp. 473-496, **WP 1309 (December 2020).**

CARRIERO A., F. CORSELLO and M. MARCELLINO, *The global component of inflation volatility,* Journal of Applied Econometrics, v. 37, 4, pp. 700-721, **WP 1170 (May 2018).**

CIAPANNA E. and G. ROVIGATTI, *The grocery trolley race in times of Covid-19. Evidence from Italy,* Italian Economic Journal / Rivista italiana degli economisti, v. 8, 2, pp. 471-498, **WP 1341 (June 2021).**

CONTI A. M., A. NOBILI and F. M. SIGNORETTI, *Bank capital requirement shocks: a narrative perspective,* European Economic Review, v.151, Article 104254, **WP 1199 (November 2018).**

FAIELLA I. and A. MISTRETTA, *The net zero challenge for firms' competitiveness,* Environmental and Resource Economics, v. 83, pp. 85-113, **WP 1259 (February 2020).**

FERRIANI F. and G. VERONESE, *Hedging and investment trade-offs in the U.S. oil industry,* Energy Economics, v. 106, Article 105736, **WP 1211 (March 2019).**

GUISO L., A. POZZI, A. TSOY, L. GAMBACORTA and P. E. MISTRULLI, *The cost of steering in financial markets: evidence from the mortgage market,* Journal of Financial Economics, v.143, 3, pp. 1209-1226, **WP 1252 (December 2019).**

LAMORGESE A. and D. PELLEGRINO, *Loss aversion in housing appraisal: evidence from Italian homeowners,* Journal of Housing Economics, v. 56, Article 101826, **WP 1248 (November 2019).**

LI F., T. MÄKINEN, A. MERCATANTI and A. SILVESTRINI, *Causal analysis of central bank holdings of corporate bonds under interference,* Economic Modelling, v.113, Article 105873, **WP 1300 (November 2020).**

LOBERTO M, A. LUCIANI and M. PANGALLO, *What do online listings tell us about the housing market?,* International Journal of Central Banking, v. 18, 4, pp. 325-377, **WP 1171 (April 2018).**

MIRENDA L., M. SAURO and L. RIZZICA, *The economic effects of mafia: firm level evidence,* American Economic Review, vol. 112, 8, pp. 2748-2773, **WP 1235 (October 2019).**

MOCETTI S., G. ROMA and E. RUBOLINO, *Knocking on parents' doors: regulation and intergenerational mobility,* Journal of Human Resources, v. 57, 2, pp. 525-554, **WP 1182 (July 2018).**

PERICOLI M. and M. TABOGA, *Nearly exact Bayesian estimation of non-linear no-arbitrage term-structure models,* Journal of Financial Econometrics, v. 20, 5, pp. 807-838, **WP 1189 (September 2018).**

ROSSI P. and D. SCALISE, *Financial development and growth in European regions,* Journal of Regional Science, v. 62, 2, pp. 389-411, **WP 1246 (November 2019).**

SCHIVARDI F., E. SETTE and G. TABELLINI, *Credit misallocation during the European financial crisis,* Economic Journal, v. 132, 641, pp. 391-423, **WP 1139 (September 2017).**

TABOGA M., *Cross-country differences in the size of venture capital financing rounds: a machine learning approach,* Empirical Economics, v. 62, 3, pp. 991-1012, **WP 1243 (November 2019).**

2023

APRIGLIANO V., S. EMILIOZZI, G. GUAITOLI, A. LUCIANI, J. MARCUCCI and L. MONTEFORTE, *The power of text-based indicators in forecasting Italian economic activity,* International Journal of Forecasting, v. 39, 2, pp. 791-808, **WP 1321 (March 2021).**

BARTOCCI A., A. NOTARPIETRO and M. PISANI, *Non-standard monetary policy measures in non-normal times,* International Finance, v. 26, 1, pp. 19-35, **WP 1251 (November 2019).**

CAPPELLETTI G. and P. E. MISTRULLI, *The role of credit lines and multiple lending in financial contagion and systemic events,* Journal of Financial Stability, v. 67, Article 101141, **WP 1123 (June 2017).**

CECI D. and A. SILVESTRINI, *Nowcasting the state of the Italian economy: the role of financial markets,* Journal of Forecasting, v. 42, 7, pp. 1569-1593, **WP 1362 (February 2022).**

CIAPANNA E, S. MOCETTI and A. NOTARPIETRO, *The macroeconomic effects of structural reforms: an empirical and model-based approach,* Economic Policy, v. 38, 114, pp. 243-285, **WP 1303 (November 2020).**

DAURICH D, S. DI ADDARIO and R. SAGGIO, *The macroeconomic effects of structural reforms: an empirical and model-based approach,* Review of Economic Studies, v. 90, 6, pp. 2880–2942, **WP 1390 (November 2022).**

DI ADDARIO S., P. KLINE, R. SAGGIO and M. SØLVSTEN, *The effects of partial employment protection reforms: evidence from Italy,* Journal of Econometrics,v. 233, 2, pp. 340-374, **WP 1374 (June 2022).**

FERRARI A. and V. NISPI LANDI, *Toward a green economy: the role of central bank's asset purchases,* International Journal of Central Banking, v. 19, 5, pp. 287-340, **WP 1358 (February 2022).**

FERRIANI F., *Issuing bonds during the Covid-19 pandemic: was there an ESG premium?,* International Review of Financial Analysis, v. 88, Article 102653, **WP 1392 (November 2022).**

GIORDANO C., *Revisiting the real exchange rate misalignment-economic growth nexus via the across-sector misallocation channel,* Review of International Economics, v. 31, 4, pp. 1329-1384, **WP 1385 (October 2022).**

GUGLIELMINETTI E., M. LOBERTO and A. MISTRETTA, *The impact of COVID-19 on the European short-term rental market,* Empirica, v. 50, 3, pp. 585-623, **WP 1379 (July 2022).**

LILLA F., *Volatility bursts: a discrete-time option model with multiple volatility components,* Journal of Financial Econometrics, v. 21, 3, pp. 678-713, **WP 1336 (June 2021).**

LOBERTO M., *Foreclosures and house prices,* Italian Economic Journal / Rivista italiana degli economisti, v. 9, 1, pp. 397-424, **WP 1325 (March 2021).**

LOMBARDI M. J., M. RIGGI and E. VIVIANO, *Worker's bargaining power and the Phillips curve: a micro-macro analysis, and wages,* Journal of the European Economic Association, v. 21, 5, pp. 1905–1943, **WP 1302 (November 2020).**

NERI S., *Long-term inflation expectations and monetary policy in the Euro Area before the pandemic*, European Economic Review, v. 154, Article 104426, **WP 1357 (December 2021).**

ORAME A., *Bank lending and the European debt crisis: evidence from a new survey*, International Journal of Central Banking, v. 19, 1, pp. 243-300, **WP 1279 (June 2020).**

RIZZICA L., G. ROMA and G. ROVIGATTI, *The effects of shop opening hours deregulation: evidence from Italy,* The Journal of Law and Economics, v. 66, 1, pp. 21-52, **WP 1281 (June 2020).**

TANZI G. M., *Scars of youth non-employment and labour market conditions,* Italian Economic Journal / Rivista italiana degli economisti, v. 9, 2, pp. 475-499, **WP 1312 (December 2020).**

2024

MORO A. and V. NISPI LANDI, *The external financial spillovers of CBDCs,* Journal of Economic Dynamics and Control, v. 159, Article 104801, **WP 1416 (July 2023).**

FORTHCOMING

BALTRUNAITE A., M. CANNELLA, S. MOCETTI and G. ROMA, *Board composition and performance of state-owned enterprises: quasi experimental evidence,* The Journal of Law, Economics, and Organization, **WP 1328 (April 2021).**

BUONO I, F. CORNELI and E. DI STEFANO, *Capital inflows to emerging countries and their sensitivity to the global financial cycle,* International Finance, **WP 1262 (February 2020).**

CORNELI F., *Sovereign debt maturity structure and its costs,* International Tax and Public Finance, **WP 1196 (November 2018).**

CUCINIELLO V. and N. DI IASIO, *Determinants of the credit cycle: a flow analysis of the extensive margin,* Journal of Money, Credit and Banking, **WP 1266 (March 2020).**

FERRARI A. and V. NISPI LANDI, *Whatever it takes to save the planet? Central banks and unconventional green policy,* Macroeconomic Dynamics, **WP 1320 (February 2021).**

FLACCADORO M., *Exchange rate pass-through in small, open, commodity-exporting economies: lessons from Canada,* Journal of International Economics, **WP 1365 (April 2022).**

GAUTIER E., C. CONFLITTI, R. FABER, B. FABO, L. FADEJEVA, V. JOUVANCEAU, J.-O. MENZ, T. MESSNER, P. PETROULAS, P. ROLDAN-BLANCO, F. RUMLER, S. SANTORO, E. WIELAND and H. ZIMMER, *New facts on consumer price rigidity in the euro area,* American Economic Journal: Macroeconomics, **WP 1375 (July 2022).**

MICHELANGELI V. and E. VIVIANO, *Can internet banking affect households' participation in financial markets and financial awarness?,* Journal of Money, Credit and Banking, **WP 1329 (April 2021).**

MISTRETTA A., *Synchronization vs transmission: the effect of the German slowdown on the Italian business cycle,* International Journal of Central Banking, **WP 1346 (October 2021).**

RAINONE E., *Reservation rates in interbank money markets,* Journal of Money, Credit and Banking, **WP 1160 (February 2021).**

RAINONE E., *Real-time identification and high frequency analysis of deposits outflows,* Journal of Financial Econometrics, **WP 1319 (December 2017).**