

# Diversity of backgrounds and ideas: the case of research evaluation of economics in Italy

Marcella Corsi, Carlo D'Ippoliti and Giulia Zacchia\*

The paper contributes to the literature on impact of research evaluation in terms of preserving and reproducing inequality and specifically gender segregation within the academia. Through a large-scale natural experiment encompassing two entire cohorts of Italian economists, we document how candidates for academic positions in economics, especially top-rank positions, are pushed to increasingly conform to a standardized research profile. We find evidence of bias in research evaluation against women (gender glass ceiling) and a substantial heterogeneity in the chances to qualify for an academic position, depending on candidates' main fields, topics and methods of research. Our results also indicate that economists working on peripheral topics (i.e. Italian economy), less popular research fields and/or with heterodox methods were less likely to qualify for top-rank academic positions, independently of their bibliometric indicators.

**KEYWORDS:** research evaluation, bibliometrics, diversity in economics.

**JEL Classifications:** J16; A14

## 1. Context and motivation

Research evaluation systems at the level of papers, journals, authors or institutions, are increasingly scrutinized for their possible contribution to preserving and reproducing inequality and specifically gender segregation within the academia (Probert, 2005). For example, at the European level the ERC and Marie Skłodowska-Curie actions are now raising the awareness of their evaluators on the risk of implicit or unconscious biases in research evaluation (European Commission, 2017; Leru, 2018). As reported by the European Commission (2012) the main concern is that “commitment to excellence and to objectivity that is a hallmark of academic life can make it particularly difficult for research institutions to recognize the ways in which standard practices may give advantage to some and disadvantage others” (p. 6).

At the same time, researchers continue to investigate the procedures of research evaluation, especially with the debate on the pros and cons of peer review versus bibliometrics, for possible biases, which could impact on the direction and development of science and the incentives for researchers and institutions. Special emphasis is devoted to the potential discouragement of radically new path-breaking ideas (Wang *et al.*, 2017), multidisciplinary methods (Rafols *et al.*, 2012; Hicks *et al.*, 2015), and research orientations and fields pursued by a minority of researchers in the respective disciplines (López-Piñero and Hicks, 2015; Kapeller and Steinerberger, 2016).

Some authors imply that women and other minorities hold specific views, and thus increasing diversity of backgrounds would automatically bring about higher diversity of scientific perspectives too (e.g. Forget, 1995; May *et al.* 2018). Others consider this to be an essentialist argument and object that “‘female’ is as ‘female’ does” (Christensen, 2001, p. 108). However, for the most part the two debates seem to proceed in parallel, with little interaction between them.

In this work we aim to contribute to bridging the two debates, analysing the impact of a specific research evaluation system on both diversity of backgrounds, and diversity of methods and approaches. We focus on the evaluation of individuals, and analyse the case of economics in Italy. This case is especially interesting for both the size (which involved a whole generation of academics in the country) and the transparency of a recently introduced centralized national procedure for the recruitment and promotion of academics: the Abilitazione Scientifica Nazionale (ASN). Within it, the case of economics has already spurred a literature due to its emphasis on bibliometrics despite the governmental inclusion of the discipline among the “non-bibliometric fields” (De Paola and Scoppa, 2015; Corsi *et al.*, 2018; Bagues *et al.*, 2017).

However, economics is an interesting case study per se: on the one hand, because it exhibits marked inequality (e.g. of salaries, prestige and citations) among academics and distinctly higher gender imbalances than many other disciplines, including the STEMs (Fourcade *et al.*, 2015; Bayern and Rose, 2016); on the other hand, because it exerts a strong influence on policymaking even beyond economic policy, and therefore its scientific debate has immediate social significance (European Commission, 2016).

---

\* Sapienza University of Rome: [marcella.corsi@uniroma1.it](mailto:marcella.corsi@uniroma1.it), [carlo.dippoliti@uniroma1.it](mailto:carlo.dippoliti@uniroma1.it), [giulia.zacchia@uniroma1.it](mailto:giulia.zacchia@uniroma1.it). We wish to thank Daniele Vannella and Liz Braunstein, Associate Director at EconLit, for precious assistance in data collection, and Lisa Cook, Orsola Costantini, Joyce Jacobsen, Annalisa Rosselli, Mauro Sylos Labini, and the participants and discussants in the IAFFE panel at the ASSA 2017 and the SIE 2016 annual meetings, as well as three anonymous referees, for helpful comments. Financial support from the Institute for New Economic Thinking is gratefully acknowledged, INET grant # INO1500040.

Indeed, the economists' infamous failure to predict the financial crisis of 2007-2008 and the ensuing Great Recession has given new impulse to advocates for more diversity in the economics profession. In a very explicit report, the Independent Evaluation Office (IEO, 2011) of the International Monetary Fund (IMF) denounced how an environment that is prone to group thinking and discourages diversity and "contrarian views" led the Fund to deemphasize macroeconomic and financial risks up until the crisis. Attached to the report are both a statement by the then managing director of the IMF and an IMF staff response (as well as a "Chairman's Summing Up"). The staff response objected that the "recommendation to increase financial expertise and staff diversity [...] does not follow from the pre-crisis experience: the vast majority of financial experts, from a diversity of countries and of backgrounds, also failed to see the crisis coming." (*ibid.*, p. 42). Interestingly, this discordance may be due not just to the legitimate willingness on the side of the staff to defend its members, but also from the fact that the IEO (and the then IMF managing director) referred to both diversity of views/ideas and backgrounds, while the staff focused on the latter only.

This narrower perspective, focused on diversity of demographic backgrounds, dominated the subsequent debate.<sup>1</sup> A growing literature, summarized by Bayer and Rouse (2016), focuses on gender diversity in particular, while sociologists and historians of economics focus on geography and nationality by looking at the institutional dominance of a handful of US universities (Fourcade *et al.*, 2015).

In a separate debate concerning diversity of ideas, single top-ranked economists have occasionally voiced (even deep) concern about the current lack of critical thinking in economic research (e.g. Krugman, 2009; Solow, 2010; or more recently Blanchard, 2016; Romer, 2016; and Rubinstein, 2017). However, these criticisms have so far remained individual acts, and have not produced a widespread debate on the usefulness and extent of pluralism in economic research.<sup>2</sup> The latter has rather remained a debate confined to the niche of self-defined 'heterodox' or 'eclectic' economists (Lee, 2006; Corsi *et al.*, 2010, 2011). Therefore, while we chose to focus on a single discipline in order to consider in-depth the contents of each researcher's publications, economics seems to be an especially suited case to highlight the need to bring together the two debates, that on background diversity and that on diversity of ideas.

## 2. Diversity within economics

Drawing from the diversity management literature, several scholars argue that department organization, research activities, criteria of evaluation and career paths are all sources and manifestations of inequality within the academia (Acker, 1990; Johansson and Śliwa, 2014). From a human resources perspective, the main object of analysis when investigating diversity is how can universities create a working environment that fosters and benefits from the creativity and productivity of diverse research groups.

Extant literature on diversity within economics is characterized by two parallel approaches. One strand of literature analyses the contributions to the discipline of economists from various backgrounds (sex, race, age, sexual orientation, etc., which we summarize here as 'background diversity'); the other focuses on social and institutional constraints to equal opportunities for economists of various backgrounds and/or various theoretical or methodological orientations ('diversity of ideas'), especially investigating the organization of human resources within research centres. We review here both strands by considering background diversity in the next section, and diversity of ideas in the following one.

### 2.1. Diversity of backgrounds

A large literature has investigated various gaps in access and career progression in universities, studying the main factors affecting those differences: productivity, in terms of both the number and type of publications; research field preferences, leading to horizontal segregation; time allocation at work, with the unequal

---

<sup>1</sup> See for example the opening speech by Janet Yellen at the *National Summit on Diversity in the Economics Profession*, Federal Reserve Board, Washington (DC), 30<sup>th</sup> October 2014.

<sup>2</sup> Indeed, several economists deny that the 2007-8 crisis is any sign of a failure of mainstream economics (Lucas, 2009), and many reject the notion that more pluralism in economic research should be sought or expected in the future: for example Rodrik (2016) notes that "[p]luralism with respect to conclusions is one thing; pluralism with respect to methods is something else. No academic discipline is permissive of approaches that diverge too much from prevailing practices" (p. 199). In contrast, in various countries pluralism in teaching is widely discussed both by academics and the media, e.g. following the expansion of the "Rethinking Economics" movement.

distribution of teaching and bureaucratic tasks; and family responsibilities. By necessity, our analysis here focuses on gender and on age inequality. In terms of age, many studies (Abramo *et al.* 2016 and van Arensbergen *et al.*, 2012) concentrated on productivity, highlighting a progress reduction of the productivity gap between men and women for the younger generation. There is evidence that other dimensions are relevant as well, for example Frank (2006) documents that in six UK universities gay and bisexual men suffer from a glass ceiling comparable to that faced by heterosexual women; moreover, Skachkova (2007) reports how, in US, being an immigrant women means still to encounter obstacles in teaching, research, interact with native colleagues, and in developing academic career, because of the intersectional effects of national origin, race, ethnicity and sex. However, these and other dimensions of diversity of backgrounds had to be neglected here because it was not possible to obtain more personal information regarding the candidates to the ASN.

In Italy, in 2017 women represented 37.4% of academic staff, and they appear to suffer from significant glass ceilings.<sup>3</sup> In its periodic report *She Figures*, the European Commission proposes the measurement of a Glass Ceiling Index (GCI), computed as a ratio of the proportion of women in academia and the proportion of women in top academic positions (full professors).<sup>4</sup> In Italy, in 2017 the GCI was 1.63 in academia generally, with values e.g. of 1.27 in the humanities and arts and 1.64 in mathematics and computer sciences. In economics, the GCI was 1.93 in 2017, compared to 2.55 in 2000. Thus, there has been some progress towards reducing the glass ceiling, although women continue to be even less represented in full professorship in economics than in academia generally, and even more so than in traditionally male-dominated fields such as mathematics. Indeed, in most countries economics exhibits the highest gender gaps in tenure and promotion rates, average salaries and job satisfaction among both the social sciences (Ginther, Kahn, 2006) and math-intensive fields (Ceci *et al.*, 2014). Reviewing this evidence, Bayer and Rouse (2016), conclude that “the field of economics is behind others in its progress on diversity concerns” (p. 238).

Within the literature on gender diversity, a recent topic of debate concerns the gender impact of quantitative (bibliometric) research assessment methods on imbalances within academia. There is ample evidence that in quantitative terms men have been outperforming women in terms of research output. For example, Baccini *et al.* (2014) find that among the reasons for this lower productivity is a higher burden of teaching duties and administrative tasks. Zacchia (2017) further documents how outside of the workplace women academics are not insulated from the more general higher burden of unpaid care and housework than men. For these reasons, some authors ask whether the equal application of bibliometric indexes may produce indirect discrimination, and argue for the creation of gender-specific rankings aimed at providing a more nuanced representation of the relative performance of men and women researchers (Abramo *et al.*, 2015).

In contrast, analysing undisclosed non-public data from a research quality assessment of Italian universities, which included a random sample of papers evaluated using both bibliometric indicators and peer review, Jappelli *et al.* (2017) argue that bibliometric evaluation does not penalize women with respect to men, whereas peer review might. This last piece of evidence confirms the results of an emerging literature on the bias of peer review (Bornmann *et al.*, 2008).

However, the literature is less clear-cut on the supposedly unbiased nature of bibliometrics. From this point of view, the ASN data used in the present work have been used by several researchers as a natural experiment. For example, De Paola *et al.* (2016) find that the introduction in 2012 of the new ASN recruitment and promotion system (described in the following section, and that relies heavily on the use of bibliometric indicators) has not increased women’s probability of being promoted either to associate or full professorship. Many studies focused on the impact of the composition of the evaluation committee on men’s and women’s chances of promotion. The results have, however, not been mutually consistent: Abramo *et al.*

---

<sup>3</sup> In economics, the share of women was even lower (30.4%); 15.7% of full professors in economics were women, 33.2% of associate professors, and 44.8% of researchers. As a consequence, for women there was a classic pyramidal structure of employment, with 19% of women employed as full professors at the top, followed by the associate professors (36%), and a larger number of researchers at the base of the pyramid (46%). By contrast, for men the hierarchical structure takes on the shape of a inverse pyramid, with the largest share represented by full professors (44%), followed by associate professors (31%), and finally by researchers (25%). For all figures in this section, the source of the data is the website of the Ministry of Research and University updated to 12/31/2017, available at: <http://cercauniversita.cineca.it/php5/docenti/cerca.php>. For economics, we consider the disciplinary field formally known as “Political economy, SECS-P01”. However, researchers that in an international context may be considered as economists may be employed in Italian universities within a different subfield of the social sciences, e.g. public policy, econometrics, etc. We are bound to ignore them here, due to the impossibility of objectively identifying the “economists” from the others, in the other research fields.

<sup>4</sup> Thus, the GCI ranges between 0 and infinity: a GCI of 1 indicates that there is no difference between women and men in terms of their chances of being promoted; and a GCI greater than 1 indicates that women are less represented in full professorship positions than in academia generally (the “glass ceiling” effect).

(2016) and Bagues *et al.* (2017) find no empirical support for the thesis that the presence of women in the evaluation committee would positively affect the chances of success of women candidates. If anything, Bagues *et al.* (2017), who also consider the Spanish case, find that in many cases gender-mixed committees tend to be less favourable towards women candidates than all-male committees. In contrast, Checchi *et al.* (2015) and De Paola and Scoppa (2015) find a significantly positive role for the presence of women in the commission on women candidates' chances of promotion. Bagues *et al.* (2017) further note that personal connections with the commission members remain a significant predictor of a candidate's chances.

In conclusion, extant literature suggests that women and other minorities may be subject to discrimination leading to lower chances of hiring and promotion. These trends lead to lower presence in academia and higher concentration in the lower ranks; episodes of discrimination have been documented within peer review too. However, mixed results have been found with respect to the extent of potential indirect discrimination arising from bibliometric methods of research evaluation.

## 2.2. Diversity of ideas

With respect to the field of economics, research evaluation has been debated less often for its impact on background diversity than for the impact on diversity of views and methods. Several studies focused on Anglo-Saxon countries (e.g. for the UK: Lee 2006, Lee *et al.* 2013; Harley and Lee, 1997; for the USA: Fourcade *et al.*, 2015; for Australia: Bloch, 2010), on German-speaking countries (Grimm *et al.*, 2018), France (Chavance and Labrousse, 2016), and Italy (Corsi *et al.*, 2010, 2011; Baccini, 2016). These works stress how research evaluation assessments, usually based on the identification of quality with citation impact, tend to create or reinforce the normative standard of a supposedly unequivocal 'right way' of doing research, even in a contested discipline such as economics, characterized by competing scientific paradigms. These findings reflect what has been noted for all disciplines (Wilsdon, 2006), that is, national research assessments tend to reward the approaches that are more visible and popular within a discipline, reinforcing the pre-existing journals rankings and more in general the discipline's hierarchy. In the case of economics, this appears to imply lower chances of hiring and promotion for so-called non-mainstream economists (e.g. in France AFEP, 2009). For example, analyzing the consequences of various rounds of the Research Assessment Exercise in the UK, Lee *et al.* (2013) find a progressive decline in the variability in approaches and pluralism within economics, especially in terms of: (i) gradual elimination of non-mainstream economists from the economics departments; (ii) concentration of economic research in five core areas (macroeconomics and growth; microeconomics and game theory; econometrics and mathematical economics; finance and monetary theory; development and transition economics); (iii) institutional dominance of few elite departments at the national level. Corsi *et al.* (2010; 2011) identified a similar trend in Italy, highlighting the role of research evaluation in shaping the incentives for institutions and single researchers.

A comparison of trends across countries could be summarized by the definition of an idealized standard of excellence for economists, which could imply: pronounced *specialization* in a few research fields and/or methods; the identification of quality with *popularity* among the peers and/or few, selected "top journals" (Heckman, 2017); as a consequence of the above, a focus on the social and economic trends in few, *central* economies, notably the US one.

These trends do not harm so-called heterodox or eclectic economists only. For example, Kates (2013) highlights the negative impact of research evaluation practices in several countries on the economists' engagement in the history of economic thought, possibly because this sub-field is not based on mathematical models and rather relies on multidisciplinary approaches. Pasinetti (2006) highlighted how top economics journals are either US- or UK-based, and therefore tend to favor the study of those countries with negative consequences for the acceptance of submissions from scholars investigating other economies. Vessuri *et al.* (2014) reiterates the point with specific reference to studies on Latin America, and Alencar de Farias (2018) talks about standards imposed by "weird (Western, Educated, Industrialized, Rich and Democratic)" countries.

As mentioned above, some scholars hint at a connection between fostering diversity of backgrounds and diversity of ideas. In the case of economics, Albelda (1997), Stastny (2010) and May *et al.* (2018) find systematic differences between men's and women's research interests and economic policy preferences. However, such differences could as well be the result of men and women economists' different workplace and lifetime experiences Forget (1995). For example, Zacchia (2017) finds that a progressively more competitive environment in the last few decades induced Italian women economists to adopt a strategy of

‘homologation’ towards the same research interests of their male colleagues, in order to increase their chances to achieve a tenured academic position.

### 3. The Italian institutional context

In a growing number of countries, decisions about hiring and/or promoting a professor are increasingly homogenized along a standard concept of excellence measured by bibliometric measures of citation impact and/or predefined lists of journals. In countries like Italy, France and Spain, promotions within academia are based on national contests, giving the opportunity of examining promotion gaps and how indirect discrimination may operate when wages are fixed at the national level and national committees decide on promotions (for studies on the France case, see e.g. Bosquet *et al.*, 2018, for Spain, Bagues *et al.* 2017). Evidently, several other countries adopt nation-wide frameworks of research assessment, e.g. for the evaluation of journals, projects, departments, or even whole disciplines or geographical areas. Some of these constitute the object of nation-wide research evaluation in Italy as well, but within completely separated procedures. Therefore, we focus here on the evaluation of single researchers.

In Italy, a 2010 reform of the university system (Law 240/2010) created the ASN procedure for recruitment and promotion aims within academia. “Professore associato” is now the lower tenured rank in Italian academia, roughly corresponding to a senior lecturer or associate professor position; and “professore ordinario” is the higher rank (corresponding to full professor).<sup>5</sup> Individuals who obtain a national qualification as associate or full professors can then compete for a job at the corresponding rank at any Italian university. Accordingly, while in Spain and France obtaining a *qualification* almost automatically leads to promotion, in Italy the chances to be promoted after having obtained a qualification are much lower (to-day, 27.1% of successful candidates to full professorship and 38.2% of candidates to associate professors have been promoted or hired).

The ASN proceeded in two annual rounds in 2012 and 2013, and was then interrupted.<sup>6</sup> The reform distinguishes between two broad cultural areas: the life and natural sciences, defined by law as “bibliometric areas”, and the social sciences and humanities, the “non-bibliometric areas”. Within the ASN, different sorts of bibliometric indicators are prescribed for the evaluation of candidates in the bibliometric and non-bibliometric areas. For the non-bibliometric areas, which include economics, the bibliometric thresholds were defined in terms of numbers of publications.<sup>7</sup> Specifically, it was expected that a qualified candidate would have a higher number than the median of the then tenured economics professors in Italy, of the following:

- (i) research monographs, excluding edited volumes, published in the past 15 years (10 for candidates to associate prof.), henceforth “first bibliometric criterion”;
- (ii) journal articles and book chapters published in the past 10 years (5 for candidates to associate prof.), the second bibliometric criterion;
- (iii) journal articles in “top journals”, the so-called *A-list*, published in the past 15 years (10 for candidates to associate prof.), the third bibliometric criterion.

For all three, the candidates’ number of publications was normalized by “academic age”, i.e. the time since the first recorded publication, subtracting periods of parental leave. The drafting of the *A-list* for each discipline was delegated by the National Agency for the Evaluation of Research and the University System (Anvur) to an ad hoc working group made up of five full professors based in an Italian university. In principle, these thresholds were formally a reference point and not a mandatory minimum requirement for the commissions. According to Law 240/2010, these indicators should have been complemented by further qualitative criteria, such as being on the editorial board of scientific journals, etc. Therefore, there is both an element of bibliometric evaluation and of peer review in the ASN.

---

<sup>5</sup> Until the reform, and for a short transitory period thereafter, there was a further lower-rank tenured position, “ricercatore universitario”, roughly corresponding to a lecturer or assistant professor. Thus, in the sample of individuals that we consider in this work there could be candidates for obtaining a qualification as associate professor, who already had tenure at the time (we include them among the “tenured professors candidates for promotion”).

<sup>6</sup> In 2016, a new procedure was introduced, in which the same methods and criteria are adopted, but instead of yearly rounds there are now quarterly, so-called “rolling” calls for applications.

<sup>7</sup> In contrast, for the bibliometric areas these thresholds were defined in terms of normalized citation counts for both articles and journals, as recorded by the databases Web of Science™ and Scopus™.

The procedure evolved into five steps: (i) candidates applied by responding to a public call; (ii) after the deadline for applications had elapsed, the Ministry for Research and Evaluation randomly selected the five members of the commission for each discipline, from a pool of candidate commissioners (who had to be full professors in an Italian university and meet the same above-mentioned bibliometric threshold criteria); (iii) once the names of the components of the commissions were drawn and made public, candidates could decide to withdraw from the procedure.<sup>8</sup> In the meantime, the Ministry computed the three bibliometric indicators for all candidates and passed them to the commission; (iv) after the deadline for withdrawing passed, the commission analysed the CVs, publications and bibliometric indexes of all candidates who had not withdrawn; and finally (v) the Ministry published the commission's assessment of all candidates for a same discipline simultaneously.

All commissions were obliged to publicly report all material related to their work (i.e. the candidates' CVs, their three bibliometric indicators, and the commission's narrative evaluation of each candidate) on a public website.

Therefore, while candidates for promotion in France and Spain are required to make oral presentations in front of a committee, in Italy all candidates are evaluated on the basis of their CVs and a sample of self-selected publications only.

#### 4. Data and descriptive statistics

We collected the CVs, the three bibliometric indicators (as certified by the commission) and the final results of all candidates to the 2012 and 2013 rounds for economics from the ASN website. We matched this information with all publications by those candidates indexed in EconLit and Google Books.<sup>9</sup> These two databases allow us to consider the following metadata for each publication: title, abstract, keywords, and JEL codes, for EconLit; and title, abstract and keywords for Google Books. From these data we obtain the final list of variables used in our analysis, reported in table A1 in appendix.

Concerning diversity in the candidates' demographic backgrounds, data limitations constrain our analysis to the consideration of age and sex only. As shown in table 1, of the 345 candidates for qualification to full professorship in economics and 525 candidates for associate professorship in 2012 and 2013, women constituted respectively 22% and 35%.<sup>10</sup> Less than 10% of candidates decided to withdraw their application once the names of the commission members were made public.<sup>11</sup> The majority of these withdrawn candidates are women. In contrast, there are no significant differences in the candidates' age by gender, by outcome at the ASN, or by the decision to withdraw, the only significant difference being between (younger) candidates as associate and (older) as full professors.

TABLES 1 AND 2 AROUND HERE

Considering candidates' background in a wider sense, we single out the macro-area of location of their main institutional affiliation (as a proxy for candidates' birthplace, given the very low mobility of academics in Italy), separately considering whether it is outside of the country. Beside a historic legacy of North-South divide and possibly discrimination against Southerners, the information may also proxy for the candidates' participation in established academic networks at the national level (D'Ippoliti, 2017). To capture the latter, we also consider candidates' average number of co-authors, and separately the number of their co-authors among the five members of the ASN economics commission (Bagues *et al.*, 2017).

---

<sup>8</sup> Not qualifying at the ASN has the only consequence of not being eligible to be a candidate in the next year. However, it is possible that candidates incurred or perceived a reputational cost that induced many to not apply or to withdraw.

<sup>9</sup> EconLit is the database maintained by the American Economic Association (AEA), indexing a large number of journals, working paper series, Ph.D. theses, books and book chapters in economics. Entries in EconLit are catalogued according to a standardized index of research methods and topics, denoted by alphanumeric symbols called "JEL codes".

<sup>10</sup> De Paola and Scoppa (2016) analyze the determinants of applying for the ASN by considering as potential candidates all academics employed at Italian universities in the immediately lower rank. This procedure probably captures the vast majority of potential and actual candidates, but it is arbitrary to the extent that no such restriction was laid down (in fact, in our sample there are several candidates from non-university research centers, foreign countries, untenured academics, and tenured academics employed on even lower ranks, such as assistant professors candidates for full professorship).

<sup>11</sup> Such low figure may indicate that most potential candidates had already internalized the likely selection criteria and had not applied in the first place, if they thought they would not meet those standards; another possibility is that their uncertainty concerning the criteria was not reduced after the selection of the commission members.

As shown in table 1, the location of candidates' main institutional affiliation is skewed toward the North and Centre of Italy, reflecting the overall distribution of academics in the country. Roughly 20% of candidates are based in a foreign university or research centre, and a large majority of them are men.

However, no significant differences between men and women appear in the number of co-authors or in the number of connections with the ASN commission (table 2). For both, the number of co-authors within the commission is significantly higher among those who qualified as associate professor than among those who did not qualify (with no differences among candidates as full professors). The average number of co-authors seems significantly higher among successful men candidates than among the unsuccessful, while for women the difference is statistically significant only for candidates as associate professor.

Table 2 highlights that the differences between men's and women's scores in the three bibliometric criteria defined by law are substantial. Instead, comparing successful and unsuccessful candidates, the picture is more mixed. All candidates met the first criterion (on the number of books standardized by academic age), because the median number of books written by tenured Italian economists was estimated to be zero. A vast majority of candidates met the other two criteria as well, suggesting that these were widely regarded as necessary conditions for obtaining a qualification and therefore potential candidates who did not meet the thresholds by far did not apply. However, for the third bibliometric criterion, i.e. the number of articles in *A-list* journals, values are significantly higher among those who qualified than among those who did not qualify at the ASN.

Concerning diversity in candidates' ideas and methods, we constructed a number of measures, aimed at proxying: (i) candidates' specialization in one main topic of research; (ii) the popularity of their research areas in the economics community; and (iii) the degree to which they work on 'peripheral' topics. Quantitative measures for all three domains were obtained by using the publications' JEL codes (an alphanumeric classification of fields of economics, created and updated by the American Economic Association), the keywords, the most frequent words among those used in the publications' metadata (title, abstract, keywords), and the candidates' most 'characterizing words'.<sup>12</sup> The latter are defined as the words with highest term frequency-inverse document frequency (TF-IDF), that is, the words that the single candidate uses most often and that the other candidates do not use frequently.<sup>13</sup>

Table 3 shows the distribution of the most frequently used JEL code by each candidate, considering macro-JEL codes, i.e. the 20 aggregates obtained by considering the first letter of each code only. As shown in the table, men tend to concentrate in the fields of microeconomics (97 candidates) and of macro and monetary economics (84), while women are more often authors in labour and demographic economics (41) and in microeconomics (31). However, no significant differences between men and women emerge in the share of successful candidates by main JEL code. Focusing on specific words and lemmas, in table 4 we report those with the largest differences in frequency between candidates who qualified at the ASN and those who did not qualify. Among the words with highest significance for economics, "model" (more often associated to mathematical or empirical works) emerges as a more frequently used lemma among those who qualified than among those who did not qualify, as do "utility" (associated with mainstream, marginalist economics), and "experiment"; while "Italy", "trade", and "public", were more often used among those who did not qualify.<sup>14</sup>

## TABLES 3 AND 4 AROUND HERE

To measure a candidate's specialization in a research field or on selected topic(s), we consider the share of the candidate's most frequently used macro-JEL code within all her publications, as well as the share of her publications that exhibit at least one of her 3 most frequently used keywords, or (separately) one of her 6

---

<sup>12</sup> Following standard practice in textual analysis, stop-words (e.g. articles, conjunctions, etc.), punctuation and apostrophes were removed in order to further improve the quality and homogeneity of data. All data were preliminarily processed with the aim of eliminating all variations due to capital letters, obvious typos, hyphenation and spacing. All compound lemmas used as official descriptors of a same JEL code, such as "monetary policy" or "economic history", were considered as a single term.

<sup>13</sup> TF-IDF is a common indicator in the information retrieval literature, obtained for each term  $i$  of each author  $j$ , as the term frequency ( $n_{ij}/d_{ij}$ , with  $n$  the number of occurrences of the term, and  $d$  the number of terms in the corpus, in our case the text obtained by merging all metadata for each candidate), divided by inverse document frequency ( $\log(D/d_i)$ , with  $D$  the total number of documents, in our case the number of candidates, and  $d_i$  the number of documents, i.e. of each candidate's merged metadata, that contain word  $i$ ).

<sup>14</sup> Further results are available from the authors upon request.

most frequently words, or one of her 6 most characterising words. Our synthetic measure of specialization is then a simple average of all these shares.

Next, we consider the frequency within EconLit of the candidate's main JEL code and separately of the her most frequent and most characterising words and keywords. By considering all publications in EconLit in the decade prior to the ASN round, we interpret these frequencies as measures of the popularity of respectively the candidate's main field of research (JEL code) and her main topics/methods (words). As a further measure of how much these fields and topics are currently fashionable, we also consider the share of the main JEL code and of the same words above, in the "top 5" economics journals (again, considering the last decade only). As Heckman (2017) documents, even beyond their bibliometric impact, the "top 5" retain a highly symbolic role of normative benchmark in the profession.

As summarized in table A1, in the estimates below we consider both a synthetic indicator (obtained by simple average) of the frequency of all words considered (3 keywords, 6 most frequent words, and 6 characterising words), and we separately consider keywords and characterising words.<sup>15</sup>

Finally, concerning the peripherality of candidates' ideas and methods, we distinguish two aspects. On the one hand, we consider the candidate's interest (or lack thereof) for the US economy and/or her interest for the Italian economy (arguably the most relevant peripheral economic area for the candidates to the Italian ASN). On the other hand, we consider a measure of the marginality of a candidate's research field by its labelling as "heterodox". For each candidate, we compute the share of heterodox publications over her total number of publications, alternatively considering as "heterodox" the publications defined by Corsi *et al.* (2018) on the basis of a list of journals and JEL codes, or those included by Cronin and Lee (2010) in his ranking of heterodox and pluralist journals. Since Lee's journal ranking defines pluralist journals those publishing both heterodox and mainstream economic contributions, the latter definition is more extensive and possibly less precise than the former.

## 5. Does breaking the glass ceiling imply homologation?

The variables listed in table A1 are used to assess the determinants of a candidate's success in the ASN by running separate probit regressions on the probability of qualifying as associate and as full professor. In order to allow for the possibility that the candidates who withdrew from the competition exhibit some systematic characteristic, which may be correlated with the other candidates' outcome in the ASN, we run probit regressions with sample selection (Van de Ven and Van Pragg, 1981). The selection equation contains variables that were known to the candidates at the time of deciding whether to withdraw their application, namely the number and kinds of publications they authored (but not the bibliometric indicators, which were only made public at a later stage), demographic characteristics, and, for candidates employed in an Italian university, some characteristics of their main affiliation (i.e. its size and the feminilization of the economics department<sup>16</sup>).

Results are reported in tables 5 and 6 respectively for candidates as associate and full professors, while the results of the respective selection equations (with positive outcome implying staying in the competition) are reported in the Appendix, in table A4. For the sake of comparison results from simple (uncensored) probit regressions are reported in table A5.

Concerning the decision to stay in or to leave the competition, we find a number of observable correlates with staying in the competition.<sup>17</sup> As already noticed by De Paola *et al.* (2016) we too find that in several estimations age and sex exerted an impact on candidates' decision to withdraw. However, in our estimates this evidence is not robust across specifications. Specifically, among candidates as associate professors younger candidates appear to be more likely to withdraw, while among candidates as full professors the opposite is true. Among candidates as full professors, women are more likely to withdraw from the

---

<sup>15</sup> We do not separately consider the most frequent words, because their frequency is significantly correlated with the frequency of the most characterising words, making it impossible to separately include both in the estimates.

<sup>16</sup> Results for separate estimations on the sub-sample of candidates based in an Italian university are available from the authors upon request.

<sup>17</sup> On the whole, the correlation ( $\rho$ ) between the error term in the probit regression for the selection equation and that in the main probit, for the result at the ASN, is statistically significant for candidates as associate professor (table 5) but not for those as full professors (table 6). The respective likelihood-ratio tests similarly suggest that data on the former may be censored while data on the latter may not. However, a comparison with the results from simple probit estimates, reported in table A5 in appendix, suggest that the bias incurred when ignoring information about the candidates who withdrew from the competition may be not very relevant, since the results between the censored and the simple probit regressions are very similar.

competition, although this impact is not always statistically different from zero, while no significant gender effect is found among candidates as associate professors. The location of a candidate's institution was consistently relevant in the choice to withdraw, though again with some differences between candidates as full or as associate professors. In contrast to Bagues *et al.* (2017), we do not find that network connections with the commission affected the choice to withdraw from the competition.<sup>18</sup>

## TABLES 5 AND 6 AROUND HERE

Concerning the qualification results, we find similar results for candidates as associate and as full professor, concerning the bibliometric indexes and the control variables, but different results concerning the impact of diversity.

The third bibliometric criterion, i.e. the number of articles in *A-list* journals normalized by candidates' age, consistently exerts a positive impact on the probability to qualify as associate or full professor, with each article increasing the predicted probability by between 5% and 7%. Surprisingly, the opposite is found for the second bibliometric criterion: the normalized number of book chapters and journal articles is estimated to exert a negative impact on the probability to qualify, with a marginal effect around -1% for each additional article or chapter. This result could descend from a perceived trade-off, on the side of the judging commission, between the quality and the quantity of candidates' publications. However, the actual existence of such trade-off has been disputed in the literature (Alencar de Farias, 2018). The normalized number of research monographs does not seem to impact on the results of the ASN (except for candidates as full professor, when not controlling for background diversity). Therefore, even though all three bibliometric criteria were prescribed by law as positive signs of a candidate's scientific maturity, it emerges that the commission privileged the indicator closer connected to candidates' visibility (the *A-list* of journals) rather than the two measuring productivity.

Finally, among the control variables the number of connections with the ASN commission turns out to be a statistically significant predictor of success only when ignoring the variables related to background or intellectual diversity. Accordingly, at least for the case of economics we cannot confirm the result by Bagues *et al.* (2017) on the relevance of personal network dynamics. In our estimates, it is much more relevant whether a candidate belongs to a demographic or an intellectual minority.

Specifically, among the variables denoting background diversity, age does not appear to have significantly affected a candidate's chances of qualifying as either associate or full professor, as did not the location of the candidate's main institutional affiliation (except for foreign affiliations, which positively contributed to the probability to qualify for candidates as full professor). However, an interesting dual pattern is found for candidates' sex. Being a woman did not *ceteris paribus* affect the chances of candidates as associate professors, but it proved to be a substantial hurdle for the qualification as full professor, with an average marginal effect between -12% and -14%. Thus, the glass ceiling observed in academic economics does not depend on lower bibliometric indicators only, but it also manifests itself as a residual, lower probability to access the highest academic rank, independently of the perceived quality of research of women and men, or of other observable characteristics (which in our estimates are respectively accounted for by the control variables and by those denoting cognitive diversity).

To further investigate this residual impact of being woman, for each candidate ( $i$ ), independently of her actual sex, we computed the difference between her predicted probability of qualifying ( $Pr$ ) assuming she was a woman ( $W$ ), and the predicted probability to qualify assuming she was a man ( $M$ ), given all her other observed characteristics ( $X$ ). To facilitate comparison, in figure 1 we report  $\Delta_i = Pr(W_i|X_i) - Pr(M_i|X_i)$  for all candidates, ordering candidates by their predicted probability to qualify. As shown in the figure, for candidates as associate professor the change in the predicted probability to qualify due to one's sex is very close to zero, along the whole distribution. However, the picture is radically different for candidates as full professors. Since the predicted probabilities are always constrained to take on values between 0 and 1, along

---

<sup>18</sup> Bagues *et al.* (2017), considering all academic disciplines, interpret their evidence as a sign that connected candidates have better information about the criteria that the commission would have applied. This hypothesis may partly explain the difference in our results. Indeed, it appears that the number of books (for candidates as full professors) and of book chapters (for all) makes a candidate less likely to withdraw from the competition, while for candidates as associate professor the number of journal articles even contributed to the decision to withdraw. It may be that candidates in economics understood the criteria that the commission was going to apply less than the candidates in other disciplines.

the predicted probability to qualify we expect to observe an impact either in a U or an inverse-U shape. This shape implies that for candidates with very high or very low predicted probabilities of qualifying, their sex did not provide a significant boost (for men) or hindrance (for women). However, for all “intermediate” candidates, for whom we estimated roughly similar probabilities of qualifying and of not qualifying, being a woman implied a significant reduction in the predicted probability to qualify as full professor, up to more than -20 percentage points.<sup>19</sup>

In conclusion, the variables summarizing background diversity do not seem to influence the outcome for candidates as associate professors, whereas at least gender does for candidates as full professor.

A less clear-cut result is found for the variables that highlight diversity of ideas, some of which appear to be relevant for candidates at both ranks, though again more so for candidates for full professorship.

Concerning candidates’ specialization in a research field or on specific topics, we find that having wide research interests, as measured by the number of different macro-JEL codes used, negatively affects a candidate’s predicted probability of qualifying as associate professor. For candidates as full professor, specialization (both in terms of JEL codes and words in metadata) seems to constitute neither a drag nor a boon.

For them, the popularity of their research interests proves much more relevant. Specifically, working in the most common JEL codes (both in terms of representation in EconLit and in the “top 5”) positively affected candidates’ probability of qualifying as full professors. In contrast, the frequency with which the candidates’ keywords, characterising words, and most frequent words appear in the metadata of other publications in EconLit appears to negatively impact on their probability to qualify as full professors. In the disaggregated estimates (columns 7 and 9 of table 6) it appears that this impact can be attributed to the most characterising words rather than to the keywords. On the whole, this evidence suggests that for candidates it may be important to display originality by saying “new” (or rather, rare) words, but still about the most popular fields.

For both candidates as associate and as full professors, working on peripheral topics significantly reduces the chances to qualify at the ASN. The share of visibly heterodox publications has a large and negative marginal effect, as does for full professors the share of heterodox and pluralist publications as defined by Cronin and Lee (2010). On average, a candidate’s use of the terms “Italy” or “USA” does not seem to have exerted a statistically significant impact on their probability to qualify for the ASN. However, given the centrality of these two variables for our argument, we analysed their possibly non-linear impact using the same method used for the gender dimension above, that is by computing the difference between a candidate’s predicted probability of qualifying if she had not used the word and the predicted probability if she did.

As shown in figures 2 and 3, for candidates for associate professorship the changes in the predicted probability of qualifying are smaller than for candidates to full professorship, and generally negligible. However, for candidates to full professorship we find working on the US economy provided a boost, by more than 10 percentage points for the “average” candidates with a predicted probability to qualify around 50%. In contrast, working on Italy proved to be a hindrance for some candidates to full professorship, reducing their predicted probability to qualify by up to more than 10 percentage points.

FIGURES 1 TO 3 AROUND HERE

In conclusion, belonging to a minority in demographic and/or scientific terms is associated to a distinctly lower chance to qualify at the ASN. It seems relevant to highlight both groups of variables, relating to background diversity and to diversity of ideas, are found to exert an impact independently of the other characteristics of the candidates, and most notably regardless their scores in the three bibliometric indicators. This finding seems in agreement with that by Powdthavee *et al.* (2017) who, in an experiment on economists in 44 universities, find that the presence of low-ranking publications in an economist’s curriculum has a damaging effect on her reputation regardless of her other publications in high-ranking journals. Similarly, in our analysis the impact of factors such as being woman, writing about Italy or being labelled as heterodox, which reduces a candidate’s prospects independently of her number of articles in *A-list* journals, could be an indication of such a negative reputational effect.

---

<sup>19</sup> The two distributions apparent in the figure, for the case of candidates to full professorship, apparently would suggest that, at the same level of predicted probability to qualify, the predicted advantage for men is not exactly equal to the predicted disadvantage for women. But in fact, this difference is insignificant, given the sample size.

## 6. Conclusions

By analysing the large-scale “national scientific qualification” procedure (ASN), we provide further evidence of the gender glass ceiling in academia as well as we document the substantial risk that, at least in the case of economics, in order to find employment and progress in an academic career, both women and men must increasingly adapt to and pursue a standardized and specific research profile.

We show that not all women suffer from discrimination to the extent of not being admitted to or not proceeding in an academic career; and not all men enjoy a status of privilege. Crucially, even after controlling for observable characteristics, a candidate’s chances to qualify as associate or full professor at the Italian ASN depended not only on her productivity and even her personal connections with the judging commission, but also on her research interests. Thus, for a fuller understanding of diversity in economics it is necessary to consider not only diversity of backgrounds, but diversity of ideas as well.

The case of the ASN, in particular in economics, is internationally interesting for its mixed application of bibliometric methods and peer review. Some scholars (Jappelli *et al.* 2017) claim that the objectivity of bibliometric indicators removes the discrimination against women. Our analysis instead shows that – since women exhibit lower bibliometric indicators on average – evaluation procedures based on equal statistical measures produce unequal results. In our sample, women candidates as associate professor have on average 2.9 *A-list* journal articles (normalized by academic age) as opposed to men’s 3.8, and candidates as full professors respectively 4.1 and 5.3. Since this indicator proved the most relevant factor shaping the commission’s decision, it is not surprising that fewer women qualified than men.

Therefore, in light of evidence that women publish fewer articles partly because of higher teaching and administrative burdens (Baccini *et al.* 2014) and that men engage more in strategic (self-)citing (King *et al.* 2017) it is relevant to ask whether the application of equal, gender-blind bibliometric rules does not in fact constitute indirect discrimination (Abramo *et al.*, 2015). This point is not limited to inequality between women and men, as the bibliometric disadvantage suffered by researchers focussed on peripheral areas and topics may justify procedures of standardization of the respective bibliometric indexes (Corsi *et al.*, 2018).

However, we also show that both women (when candidates at the top rank) and economists working on peripheral topics were less likely to qualify at the ASN independently of their bibliometric indicators. For women, this is manifest by the usual unexplained residual in the regressions, the woman dummy variable. Concerning research topics, a certain degree of originality may be praised, as it emerges from the use of “characterising words” in one’s metadata. However, our estimates imply that the ASN commission effectively policed the legitimate boundaries of academic economics by systematically less often granting the qualification to those who published on the Italian economy, on less popular research fields, and/or with heterodox methods.

Therefore, the peer review component of the procedure does not appear to constitute a remedy against the potential bias of the bibliometric indicators used. From this point of view, our analysis provides new evidence on the risks of bias in peer-review that, again, were separately highlighted concerning sexism (Marsh, 2011) and lack of pluralism (Bornmann *et al.*, 2008).

In conclusion, the Italian case for economics suggests that both peer review and bibliometric methods do not necessarily provide a remedy against the discrimination against women or the tendency of researchers to suppress dissenting views. Moreover, tackling one does not automatically address the other. Therefore, if fostering diversity of backgrounds and intellectual diversity is an aim of research policy, it should be explicitly included as an objective of research evaluation practices, with clear and binding guidelines.

## REFERENCES

- Abramo, G., Cicero, T., C.A. D’Angelo (2015), “Should the research performance of scientists be distinguished by gender?”, *Journal of Informetrics*, 9: 25–38.
- Abramo, G., D’Angelo, C.A., F. Rosati (2016), “Gender bias in academic recruitment”, *Scientometrics*, 106(1): 119–141.
- Acker, J. (1990), “Hierarchies, jobs, bodies: a theory of gendered organizations”, *Gender & Society*, 4(2): 139–58.
- Albelda, R. (1997), *Economics and Feminism: Disturbances in the Field*, New York: Twayne Publishers.
- Alencar de Farias S. (2018) The Brazilian Little Way in Academia, *BAR – Brazilian Administration Review*, 15(1), epub, available at: <http://dx.doi.org/10.1590/1807-7692bar2018180035>

- Association Française d'Economie Politique (AFEP) (2009), *Evolution of economics professors' recruitment since 2000 in France*. available at: <http://assoéconomiepolitique.org/wpcontent/uploads/FAPE-State-of-pluralism-in-France-FinalVersion.pdf>
- Baccini, A., Barabesi, L., M. Cioni (2014), "Crossing the hurdle: the determinants of individual scientific performance", *Scientometrics* (101)3: 2035–2062.
- Baccini, A., G. De Nicolao (2016), "Do they agree? Bibliometric evaluation versus informed peer review in the Italian research assessment exercise", *Scientometrics*, 2016(108): 1651-1671.
- Bayer, A., C.E. Rouse (2016), "Diversity in the Economics Profession: A New Attack on an Old Problem", *Journal of Economic Perspectives*, 30(4): 221–242.
- Bagues, M., Sylos-Labini, M., N. Zinovyeva (2017), "Does the Gender Composition of Scientific Committees Matter?", *American Economic Review*, doi: 10.1257/aer.20151211.
- Blanchard, O. (2016), "Do DSGE Models Have a Future?", *Peterson Institute of International Economics*, PB 16-11.
- Bloch, H. (2010), "Research Evaluation Down Under: An Outsider's View from the Inside of the Australian Approach", *American Journal of Economics and Sociology*, 69 (5): 1530–1552.
- Bornmann, L., Rüdiger, M., D. Hans-Dieter (2008), "How to detect indications of potential sources of bias in peer review: A generalized latent variable modelling approach exemplified by a gender study", *Journal of Informetrics*, 2 (2008): 280–287
- Bosquet, C., Combes, P.P., C. García-Peñalosa (2018), "Gender and Promotions: Evidence from Academic Economists in France", *The Scandinavian Journal of Economics*, doi: 10.1111/sjoe.12300
- Ceci, S.J., Ginther, D.K., Kahn, S., W.M. Williams (2014), "Women in Academic Science: A Changing Landscape", *Psychological Science in the Public Interest*, 15(3): 75-141.
- Cecchi, D., Cicognani, S., N. Kulic (2015), "Gender quotas or girls' networks? Towards an understanding of recruitment in the research profession in Italy", *Alma Mater Studiorum-Università di Bologna*, WP DSE 1047.
- Christensen K. (2001), "'Thank God ... I Thought for a Moment You Were Going to Confess to Converting to Socialism!': Gender and Identity in Deirdre McCloskey's Crossing", *Feminist Economics*, 7(2): 105-120.
- Corsi, M., D'Ippoliti, C., F. Lucidi (2010), "Pluralism at Risk? Heterodox Economic Approaches and the Evaluation of Economic Research in Italy", *American Journal of Economics and Sociology*, 69 (5): 1495–1529.
- Corsi, M., D'Ippoliti, C., F. Lucidi (2011), "On the evaluation of economic research: The case of Italy", *Economia Politica*, 28 (3): 369–402.
- Corsi, M., D'Ippoliti, C., G. Zacchia (2018), "A Case Study of Pluralism in Economics: The Heterodox Glass Ceiling in Italy", *Review of Political Economy*, doi: 10.1080/09538259.2018.1423974.
- Cronin, B.C., F.S. Lee (2010), "Research Quality Rankings of Heterodox Economic Journals in a Contested Discipline", *American Journal of Economics and Sociology*, 69(5): 1409–1452.
- De Paola, M., V. Scoppa (2015), "Gender Discrimination and Evaluators' Gender: Evidence from Italian Academia", *Economica*, 82(325): 162-188.
- De Paola, M., Ponzio, M., V. Scoppa (2016), "Are Men Given Priority for Top Jobs? Investigating the Glass Ceiling in the Italian Academia", *IZA Discussion Paper*, 9658.
- D'Ippoliti, C. (2017), "Many-citedness", *Inet WP*, n. 57, available at: [https://www.ineteconomics.org/uploads/papers/WP\\_57-DIppoliti-revised.pdf](https://www.ineteconomics.org/uploads/papers/WP_57-DIppoliti-revised.pdf).
- European Commission (2012), *Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation*, DG for Research and Innovation, EC, Brussels.
- European Commission (2016), *She Figures 2015*, doi:10.2777/064694.
- European Commission (2017), "Implicit gender biases during evaluations: How to raise awareness and change attitudes?", *Workshop report*.
- Forget, E. (1995), "America women economists, 1900-1940: doctoral dissertation and research specialization.", in Dimand, M.A., Dimand, R., E. Forget (Eds.), *Women of value*: 25-38, Edward Elgar.
- Fourcade, M., Ollion, E., Y. Algan (2015), "The Superiority of Economists", *Journal of Economic Perspectives*, 29(1): 89–114.
- Frank J. (2006), "Gay Glass Ceilings", *Economica*, 73: 485–508.
- Ginther, D.K., S. Khan (2006), "Does Science Promote Women? Evidence from Academia 1973-2001", *NBER Working Paper*, n. 12691, Cambridge (MA).
- Grimm, C., Puhlinger, S., J. Kapeller (2018), "Paradigms and Policies: The state of economics in the german-speaking countries", *ICAE Working Paper*, 77.
- Harley, S., F.S. Lee, (1997), "Research selectivity, managerialism, and the academic labour process: the future of nonmainstream economics in U.K. universities", *Human Relations*, 50: 1425-1460.
- Heckman, J.J (2017), "Publishing and Promotion in Economics: The Curse of the Top Five", *2017 INET Plenary Conference*.
- Hicks, D., Wouters, M., Waltman, L., De Rijcke, S., I. Rafols (2015), "Bibliometrics: the Leiden Manifesto for research metrics", *Nature* 520: 429-431.
- Independent Evaluation Office (IEO) (2011), "IMF Performance in the Run-Up to the Financial and Economic Crisis. IMF Surveillance in 2004–07", *Evaluation Report*, Washington (DC): International Monetary Fund.
- Jappelli, T., Nappi, C.A., R. Torrini (2017), "Gender Effects in Research Evaluation", *Research Policy*, 46(5): 911-924.

- Johansson, M., M. Śliwa (2014), "Gender, Foreignness and Academia: An Intersectional Analysis of the Experiences of Foreign Women Academics in UK Business Schools", *Gender, Work & Organization*, 21(1): 18-36.
- Kapeller, J., S. Steinerberger (2016), "Emergent phenomena in scientific publishing: A simulation exercise", *Research Policy*, 45(10): 1945-1952.
- Kates S. (2013), *Defending the History of Economic Thought*, Cheltenham, UK: Edward Elgar.
- King, M.M., Bergstrom, C.T., Correll, S.J., Jacquet, J., J.D. West (2017), "Men Set Their Own Cites High: Gender and Self-citation across Fields and over Time", *Socius: Sociological Research for a Dynamic World*, 3: 1 -22.
- Krugman, P. (2009), "How Did Economists Get It So Wrong?", *The New York Times Magazine*, Sept. 2<sup>nd</sup>.
- Lee, F.S. (2006), "The Research Assessment Exercise, the state and the dominance of mainstream economics in British universities", *Cambridge Journal of Economics*, 31(2): 309–325.
- Lee, F.S., Pham, X., G. Gu (2013), "The UK Research Assessment Exercise and the Narrowing of UK Economics", *Cambridge Journal of Economics*, 37(4): 693–717.
- Leru (2018), "Implicit bias in academia: A challenge to the meritocratic principle and to women's careers – And what to do about it", *Leru Advice Paper*, 23.
- López-Piñero, C., D. Hicks (2015), "Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation", *Research Evaluation*, 24 (1): 78-89.
- Lucas, R.E. (2009), "In defense of the dismal science", *The Economist*, Aug. 6<sup>th</sup>.
- May, M.A., McGarvey, M., D. Kucera (2018), "Gender and European Economic Policy: A Survey of the Views of European Economists on Contemporary Economic Policy", *Kyklos*, 71(1): 162-183
- Marsh, H.W., Jayasinghe, U.W., N.W. Bond; (2011), "Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model.", *Journal of Informetrics*, 5(2011): 167-180.
- Pasinetti, L.L. (2006), "Appendix 4. A Note on Points of Dissent", *CIVR Final Report for Area 13*, Rome: Anvur, available at: [http://vtr2006.cineca.it/pubblicazioni/Area\\_13.pdf](http://vtr2006.cineca.it/pubblicazioni/Area_13.pdf).
- Powdthavee, N., Riyanto, Y.E., J.L. Knetsch (2017), "Impact of Lower Rated Journals on Economists' Judgments of Publication Lists: Evidence from a Survey Experiment", *IZA Discussion Papers*, n. 10752.
- Probert, B. (2005), "I just couldn't fit it in": Gender and unequal outcomes in academic careers", *Gender, Work & Organization*, 12(1): 50-72.
- Rafols, I., Leydesdorff, L., O'Hare, A., P. Nightingale (2012), "How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management", *Research Policy*, 41(7): 1262-1282
- Rodrik, D. (2016), *Economics Rules*, New York: W.W. Norton.
- Romer, P. (2016), "The Trouble with Macroeconomics." Mimeo, Stern School of Business, New York University, 14th September, 2016. (Forthcoming in *The American Economist*)
- Rubinstein, A. (2017), "Comments on Economic Models, Economics, and Economists: Remarks on *Economics Rules* by Dani Rodrik", *Journal of Economic Literature*, 55(1): 162–172.
- Skachkova, P. (2007), "Academic careers of immigrant women professors in the U.S.", *Higher Education*, 53(6): 697-738.
- Solow, R. (2010), "Building a Science of Economics for the Real World", Statement Prepared for Congressional Testimony before the House Committee on Science and Technology's Subcommittee on Investigations and Oversight, 20 July, Washington (DC).
- Stastny, D. (2010), "Czech Economists on Economic Policy: A Survey", *Econ Journal Watch*, 7(3): 275-87.
- van Arensbergen, P., van der Weijden, I., P. van den Besselaar (2012), "Gender differences in scientific productivity: A persisting phenomenon?", *Scientometrics*, 93: 857–868.
- Van de Ven, W. P., B. M. Van Praag (1981), "The demand for deductibles in private health insurance: A probit model with sample selection", *Journal of econometrics*, 17(2): 229-252.
- Vessuri, H., Guédon, J. C., A. M. Cetto (2014), "Excellence or quality? Impact of the current competition regime on science and scientific publishing in Latin America and its implications for development", *Current Sociology*, 62(5): 647-665.
- Wang, J., Veugelers, R., P. Stephan (2017), "Bias against novelty in science: A cautionary tale for users of bibliometric indicators", *Research Policy*, 46(8): 1416-1436.
- Wilsdon, J. (2006), *The metric tide: independent review of the role of metrics in research assessment and management.*, Sage.
- Zacchia, G. (2016), "Segregation or homologation? Gender differences in recent Italian economic thought", *MPRA Working Paper*, n. 72279.
- Zacchia, G. (2017), "Diversity in Economics: A Gender Analysis of Italian Academic Production", Inet WP, n. 61, available at: [https://www.ineteconomics.org/uploads/papers/WP\\_61-Zacchia-Diversity-Final.pdf](https://www.ineteconomics.org/uploads/papers/WP_61-Zacchia-Diversity-Final.pdf).