

The impact of a national research assessment on the publications of sociologists in Italy

Aliakbar Akbaritabar ^{1,*}, Giangiacomo Bravo ^{2,3} and Flaminio Squazzoni ⁴

¹Max Planck Institute for Demographic Research (MPIDR), Laboratory of Digital and Computational Demography, Konrad-Zuse-Str. 1, 18057, Rostock, Germany, ²Department of Social Studies, Linnaeus University, Växjö, Sweden, ³Linnaeus University Centre for Data Intensive Sciences and Applications, Linnéuniversitetet, SE-35195, Växjö, Sweden and ⁴Department of Social and Political Sciences, University of Milan, Via Conservatorio 7, 20122, Milan, Italy

*Corresponding author. E-mail: akbaritabar@demogr.mpg.de; akbaritabar@gmail.com.

Abstract

This article investigates the impact of the second national research assessment (VQR 2004–10), which was performed in 2011 by the Italian National Agency for the Evaluation of Universities and Research Institutes, on publication strategies by sociologists in Italy. We reconstructed all publications from Italian sociologists in Scopus between 2006 and 2015, that is five years before and after the assessment. We also checked academic tenure and promotions during the assessment. Our results showed the potentially distortive effect of institutional signals on publications given that Italian sociologists published more in journals that were considered influential for assessment, some, however, being of doubtful quality. Our findings would suggest that the use of informed peer review and *ad hoc* journal ranking could stimulate adaptive responses based on strategic journal targeting to ensure publication.

Key words: research assessment; ANVUR; sociologists; Italy; VQR 2004–10

1. Introduction

The dominant worldwide ‘publish or perish’ culture in academia is often associated with regular and pervasive quantitative research assessments of scientific outputs in terms of publications. The ‘meritocratic culture’ of academia requires systematic and impartial assessment as a means to legitimize funding allocation on a competitive basis (Abramo and D’Angelo 2015; Jappelli et al. 2017). These assessments are often key for promotion and career advancement (Abramo et al. 2014; Edwards and Roy 2017; Nederhof 2006).

Previous research has suggested that these assessments may have serious implications on scientists’ behaviour, though it is difficult to measure the effect of these institutional impulses (Rijcke et al. 2016; Zacharewicz et al. 2019). Here, context matters and developing comprehensive and robust indicators of evaluation are always difficult, also considering that research is rarely pursued individually (Waltman 2018). Once introduced, the performative nature of quantitative evaluation can also trigger gamification (Seeber et al. 2019; Baccini et al. 2019), which in turn risks distorting academic freedom, encouraging short-termism of research investment and portfolios (Hicks et al. 2015; Rijcke et al. 2016), even nurturing an excessively competitive environment (Edwards and Roy 2017).

In this respect, Rijcke et al. (2016) suggested that metrics-based evaluation may have four negative implications. Firstly, metrics could affect the strategic behaviour of scientists leading to switching goals with means. This occurs when evaluation itself becomes a scientist’s goal, and research activity is expected to produce publishable outcomes (Dupps and Randleman, 2012; Šupak Smolčić 2013) and maximize evaluation data/metrics. This would in turn increase scientists’ risk aversion and disincentivize long-term, challenging research projects (e.g. Butler (2003); see also van den Besselaar et al. (2017) for a critical re-examination and Hicks (2017) and Gläser (2017) for a critical response). Secondly, bibliometric indicators could embody bias against interdisciplinary research, especially when combined with peer review. These indicators could also encourage scientists to reduce task complexity and standardize collaboration (Whitley, 2007), reducing diversity of approaches, methods, and standards, while imposing rigid constraints on publication formats. Finally, these evaluation metrics may also have some effect on academic institutions, by increasing resource allocation towards more productive institutions capable of creating cumulative advantages (Abramo and D’Angelo 2015). To reconstruct the potential effect of these complex factors on scientists’ behaviour, considering research assessment in contexts where productivity, quantification

and indicators are not intrinsic to the pre-existing reward and incentive structure of academia could give useful insights.

This article aims to analyse the impact of a recent national research assessment on a sample of Italian sociologists, who are members of an academic community considered relatively unfamiliar with the ‘cult’ of productivity, quantification, and indicators (Abramo et al. 2009). The metrics for assessments here were a mix of quantitative and informed peer review, which could have further made the assessment more complex given the effect of a journal’s prestige could be mixed with reviewers’ personal criteria for quality and merit. Our analysis attempts to reveal adaptive responses by academics to ambiguous institutional signals to optimize the balance between their effort and possible outcomes in the most rational way possible, in-line with the criteria used in the national research assessment.

This article is organized as follows: Section 2 presents our research background and the motivation of our study. Section 3 presents the methodology and data, while Section 4 presents the results. Finally, Section 5 summarizes the main findings and discusses certain limitations of the study.

2. Background

Valuations, assessments and rankings are pillars of academic ‘audit culture’, and especially so in research where indicators are used to promote accountability and public governance of resources, using a mix of tools and approaches (Sauder and Espeland 2009). For instance, certain European countries have adapted different research evaluation systems, often with a combination of bibliometric indicators and peer review (Hicks 2012; Jonkers and Zacharewicz 2016; Zacharewicz et al. 2019). In a thorough analysis of performance-based research funding systems in EU28 member countries, Zacharewicz et al. (2019) found important differences in terms of evaluation objectives (e.g. only teaching activities, or a combination of research and teaching) and the temporal dimension (e.g. *ex ante* or *ex post* assessment). They suggested that any institutional design of evaluation needs to consider contexts and specificities at the expense of generalization and comparison.

For instance, Sivertsen (2016) discussed the ‘Norwegian model’, followed by various countries, such as Belgium (Flanders), Denmark, and Norway. This model has three main components: (1) a national database of all publications; (2) a single publication indicator with weights to accommodate certain field specificities and publishing traditions; and (3) a performance-based funding model that allocates resources based on the weight of institutions on the total publications at the country level. It aims to reflect on certain differences between the humanities and other fields, such as the diversity of scholarly communication means in the former, for example, national journals and books. However, Engels and Guns (2018) highlighted the difference of the Flemish adaptation of the Norwegian model, such as the use of a commercially curated database for social sciences and humanities to complement current sources (e.g. Web of Science (WoS)). As suggested by Aagaard and Schneider (2016), while concentrating on the specific case of Denmark, the evolving contexts of evaluation and the multiplicity of policies would not help reflection on the fundamental drivers of scientific performance of either institutes or academics.

The Italian case is a good example. This country has recently made progress towards attaining a viable national research-based evaluation (Hicks 2012; Jonkers and Zacharewicz 2016). After

preliminary attempts (i.e. VTR¹), in October 2006 the Italian government established a public independent agency, called the ‘Italian National Agency for the Evaluation of University and Research Institutes’ (henceforth ANVUR) (ANVUR 2006), with the institutive Law published only four years later with DPR 76/2010. The agency started to operate in February 2011 with the nomination of its Board of Directors. ANVUR’s mission was to assess research productivity of universities and research institutes. Its mandate was to establish more competitive criteria to improve the allocation of university budgets and link promotion and careers to research productivity (e.g. scientific output) and scientific merit. Besides assessing research output, it also evaluated teaching, administrative performance, social impact, and student competence by covering all areas, from hard sciences to music conservatories (Benedetto et al. 2017b). ANVUR members include well-known Italian scientists from various scientific fields appointed by Italian Ministry of University and Research (MIUR) (Ancaiani et al. 2015; ANVUR 2013; Benedetto et al. 2017b).

The first National assessment by ANVUR was VQR 2004–10, which started with a call for participation published on 7 November 2011 (basically overlapping the introduction of the national scientific qualification as a requirement for promotion and careers of all academics (Abramo and D’Angelo 2015; Marzolla 2016)). About 185,000 research products were evaluated in fourteen research areas. In STEM² and hard sciences, nearly 90 per cent of products were journal articles. In the humanities and social sciences, journal articles covered only 26 per cent of the total number of products, the rest being mostly books and book chapters. Academics working in research institutes were required to submit six research products, while those in universities with teaching duties were required to submit three such products for evaluation. An exception was considered for younger researchers based on the time of their employment in academia (Ancaiani et al. 2015; ANVUR 2013).

ANVUR selected 450 experts as ‘Groups of experts in evaluation’ (GEVs), which reflected the structure of disciplinary sectors that dominate the organization of Italian academia. They were asked to define the details and methodology of evaluation, including deciding either to use a combination of peer review, publications, and bibliometric analysis or exclusively peer review to assess the scientists’ production. Under pressure from GEVs and many academics in the humanities³, ANVUR decided to follow a mix of quantitative metrics, such as the *number of publications*, *citations* and *h-index* in hard sciences, and *qualitative metrics*, such as peer review in the humanities and social sciences. For the latter, VQR 2004–10 complied with the so-called *informed peer review*. Scientists were required to preselect and submit their *best published* research products, while VQR peer reviewers could identify each author and the impact of the journal or the prestige of the book series in which products were published via online sources. This also occurred when products were articles previously published in peer-reviewed, authoritative journals. They were peer-reviewed again by GEV experts after publication. Each product was categorized as *Excellent*, *Good*, *Acceptable* or *Limited* (Ancaiani et al. 2015; ANVUR 2013; Bertocchi et al. 2015).

After the publication of evaluation results, members of ANVUR found a certain degree of agreement between peer review and bibliometric analysis (Bertocchi et al. 2015). However, other studies re-examined the situation and found inconsistencies (Baccini et al. 2020; Baccini and De Nicolao 2016), while some concluded that ‘bibliometric indicators could be used only to assess hard sciences’ (Abramo et al. 2009). In addition, supporters of bibliometric

evaluations outlined the excessive cost of peer review (Geuna and Piolatto 2016; Zacharewicz et al. 2019), insisting on the subjective bias of peer reviewers and argued the higher fidelity of quantitative exercises that measured each scientist's productivity without adding selection distortion (Abramo and D'Angelo 2011a; Zacharewicz et al. 2019).

In one study, Bertocchi et al. (2015), who were members of the panel which evaluated research output of scientists in scientific area 13 (Economics, Management, and Statistics), selected 590 random papers from 5,681 reviewed by VQR 2004–10 experts. GEV in this field opted for *informed peer review*, while peer reviewers used bibliometric indicators and journal impact indices to assess the quality of research products. They found substantial agreement between bibliometrics and peer review. However, they also suggested a potential ambiguity of these non-blind processes, as it was impossible to establish whether close agreement was due to reviewers' opinion or to the prestige of the outlet where articles were published.

For instance, by using a subset of ANVUR data of VQR 2011–14, Traag et al. (2020) examined the respective agreement between bibliometrics and peer review in assessing the performance of academic institutions. They found that bibliometric and peer review outcomes converged the most when considering aggregate institutions rather than individual publications and when estimating the degree of internal agreement between reviewers. While outlining possible bias and pitfalls, they concluded that in some fields, metrics could be used as an alternative to peer review in research assessment.

Besides the interesting mix of quantitative metrics and criteria, Italy is worth reviewing also for other reasons. Firstly, while ANVUR assessment aimed to evaluate academic and research structures rather than individual scientists, the Minister, the agency and the press also explicitly considered the importance of productivity and quantitative indicators (Turri 2014). Secondly, ANVUR was also involved in developing common standards for the national qualification of all new associate and full professors, which linked promotion and resources to research productivity. It needs to be noted that these standards generated widespread debate within the scientific community and were opposed by many academics, sometimes generating contrasting outcomes (Baccini et al. 2020; Baccini and De Nicolao 2016). However, all these initiatives marked the beginning of a cultural shift in the institutional environment of Italian academia, making quantitative research assessments and bibliometric databases more popular, for example, *h*-index, WoS, and Scopus (Akbaritabar et al. 2018; Bonaccorsi 2018).

The case of sociologists could also give us an important insight. While the impact of ANVUR assessment has been extensively examined for the hard sciences and economics, management, and statistics (Abramo et al. 2009; Bertocchi et al. 2015), sociologists have not yet been considered. On the one hand, ANVUR followed the dominant opinion of academics when considering bibliometric indicators as being *unreliable* for assessing the productivity of social scientists (Benedetto et al. 2017a). On the other hand, sociologists are part of a community that includes not only qualitatively oriented academics, who are predominantly anti-bibliometric and might preferably publish in national journals, but also quantitative social scientists, whose research standards are closer to hard scientists, are familiar with bibliometric indicators and preferably publish in international journals. This co-existence of different epistemic communities within the same discipline could make research assessment even more problematic and worthwhile (Akbaritabar et al. 2018). Furthermore, differing sub-communities of Italian sociologists have been considered disconnected from each other, even if they have

overlapping thematic areas of focus with a drastically diverse collaboration behaviour (Akbaritabar et al. 2020). This makes them an interesting case to see how these sub-communities have reacted to ANVUR evaluation.

Another point to consider was the decision by ANVUR, under pressure from GEVs, to allow these experts to classify the quality of journals in their own field, regardless of international prestige. This added a further layer of ambiguity to the policy signal. It is worth noting here that most Italian journals with a national academic readership were considered the equal of international ones, though internal peer review procedures, impact factor, and citation rates are not comparable. This may have led rational, adaptive scientists to minimize the risk of delaying publication by targeting national outlets, which were included in the ANVUR's selected list of journals. It is also worth considering that assessment rewards and penalties were addressed to the universities and departments rather than to teams or individual scientists. This may have created incentives for less intrinsically motivated, productive scientists towards minimizing effort.

3. Data and measurements

Data were collected from Scopus in September 2016 and included all records published by Italian sociologists between 2006 and 2015. This period covered five full years before and after ANVUR, whose original call for participation was on 7 November 2011. By considering five years before and after the call, we aimed to trace pre-existing behaviour and examine scientists' reactions to institutional policies.

We started from available institutional data on the MIUR website and gathered a list of all sociologists currently registered in Italian Universities and research institutes, that is, 1,029 scholars. We reconstructed each academic's grade (i.e. assistant, associate or full professor), the 'scientific disciplinary sector'⁴ in which they were formally assigned to, gender, affiliation, department, and finally their first and last names. Promotion and careers were reconstructed by comparing data from 2010 to 2016. This was coded as a dichotomous variable in dataset: 'academic level changed' or 'unchanged'.

Regarding publication records, Scopus included approximately 8,698 journals in the social sciences (Scopus 2017). These include some of the sociology journals published by Italian publishers or edited by Italian sociologists, such as *Sociologica*, *Sociologia*, *Rassegna Italiana di Sociologia*, *Salute e Società*, *Studi Emigrazione*, *Stato e Mercato*, *Italian Sociological Review*, *Journal of Modern Italian Studies*, *Etnografia e Ricerca Qualitativa*, and *Polis*. This confirmed that popular and mainstream Italian journals were represented in the dataset.⁵ However, in VQR 2004–10, ANVUR categorized some of the Italian journals (including a total of fifty journals from which 50 per cent were indexed in Scopus (last checked on 1 August 2018) under the title of '*Fascia journals*' considering three levels, *Fascia A*, *Fascia B*, and *Fascia C* (ANVUR 2012). To complete our dataset, we then coded all articles in the sample as *Fascia* (either A, B, or C) or *non-Fascia* journals. Note that although ANVUR's list also included some prestigious journals, the list excluded many renowned international journals.

Following Akbaritabar et al. (2018), we used a productivity indicator developed by Abramo and D'Angelo (2011b) by considering a micro-economic stance to measure the scientists' output and examine the effect of certain institutional and structural factors. This function requires time for the input of academic work and publications and impact of publications as output, while also considering the number of contributors to any given scientific publication. This

index was called Fractional Scientific Strength (FSS) as Equation (1) shows:

$$FSS = \frac{1}{t} \sum_{i=1}^N \frac{c_i}{c} f_i \quad (1)$$

- t : The time window between first publication and the last one for each researcher.⁶ In case an author had one or several publications published in a single year and they did not have other publications in the sample, we considered one year as the time input.
- N : Number of sociologist's publications.
- i : Each sociologist's publication.
- c_i : Number of citations that each publication i collected.
- \bar{c} : Average number of citations of all other records published in the same year.⁷
- f_i : Inverse of the number of authors (fractional contribution of each author to paper i).

To measure the scientists' level of international collaborations (Akbaritabar et al. 2018; Katz and Martin 1997), we used an 'internationalization index', which considered the co-authors' affiliation and country (Leydesdorff et al. 2014) and calculated the number of authors with non-Italian affiliations a_{fi} on the total number of authors of each paper a_i . We aggregated this value by averaging the internationalization scores of all publications, N as follows in Equation (2):

$$InstScore = \frac{1}{N} \sum_{i=1}^N \frac{a_{fi}}{a_i} \quad (2)$$

3.1. Repeated measurement model with nested membership

In order to examine the importance of institutional embeddedness and compare individuals at different institutional levels before and after ANVUR, we followed Akbaritabar et al. (2018) in assuming that each scientist was nested in different clusters, possibly influencing their scientific productivity. We considered two cluster levels: (1) the department, as the first level of organizational embeddedness and (2) the university,⁸ relevant for power, career and strategic relationships, and sometimes also for research and collaboration (Abramo et al. 2016; Akbaritabar et al. 2020). This was to understand whether differences in these levels could be associated with different reactions to institutional policies. Sociologists in similar departments in different universities could be exposed to similar cultural contexts, and sociologists in different departments of the same university could be exposed to similar organizational facilities and constraints. In order to accommodate this complexity, we used a *nested membership random effects structure* (Baayen et al. 2008).

In order to model ANVUR impact among different departments and universities, we followed previous research (Faraway 2005; Snijders and Bosker 1999; Zuor et al. 2009) and used a *repeated measurements mixed effect model* with nested membership structure in departments and universities. This was to create 'between' (different individuals in each condition) and 'within' (same individuals before and after ANVUR) group measures that could help us understand any possible heterogeneity in individual responses. We then kept the same random effects structure for each model so that important variables were compared without excessive complications or computational inefficiencies.⁹ In case of the *total number of*

publications due to the *count* nature of this variable and to control for the robustness of our results, we ran separate models with a *negative binomial distribution* (while keeping the random effects structure similar to our other models).

To investigate the probability of publishing in 'Fascia' journals after ANVUR, we used *logistic* regression models with differing control variables. To further control the robustness of our analysis and results, we ran Bayesian models with baseline and some of the main control variables, whose results are presented in the next section.

4. Results

It is important to note that despite the vast coverage of publications in Scopus (2017), our dataset only covered a fraction of the scientific productivity of Italian sociologists (which is limited even more when focusing only on the five years before and after ANVUR). Indeed, only 57.53 per cent of the 1,029 sociologists had at least one publication record indexed in Scopus. Most missing records were probably published in national outlets, including book series by national publishers (Akbaritabar et al. 2018, 2020). We excluded all Italian sociologists without any publications from our analysis.

Figure 1 shows that the distribution of publications was highly skewed with a few sociologists publishing a considerable amount of the total number of publications. This was in line with previous studies, which showed that productivity is cumulative and non-linear (Coile 1977; Nygaard 2015; Ramsden 1994). Table 1 presents a descriptive view (mean and median) of the total number of publications and the FSS of sociologists compared over *geographical regions, universities of different ranks, sectors, departments, academic levels*, and contrasting those promoted to a new academic position from 2010 to 2016 (indicated as *level change*). It shows that if we consider the average FSS, average publications and median publications have *increased* post-ANVUR. In some cases, the median FSS has decreased post-ANVUR. It is worth noting that here we only focused on the most productive sample of Italian sociologists since those without a single publication indexed in Scopus were excluded from our dataset.

We then considered multi-level repeated measures with nested structure for each relevant variable such as the *FSS, overall citations, internationalization, number of co-authors*, and *number of*

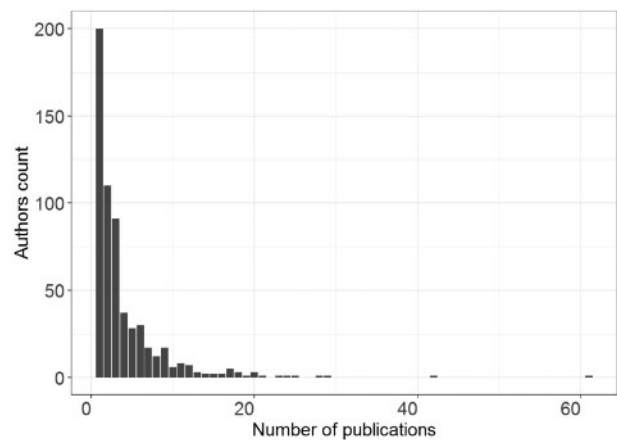


Figure 1. Distribution of the total number of publications 2006–15 (Scopus data) (X = number of publications, Y = number of authors with that many publications, 437 authors did not have any publications, which were removed from visualization and analysis).

Table 1. Descriptive statistics of the total number of publications and FSS, BF = Before ANVUR, AF = After ANVUR.

Main Category	Sub Category	Average FSS		Median FSS		Average pubs		Median pubs	
		BF	AF	BF	AF	BF	AF	BF	AF
Geo Region	Centre	0.202	0.438	0.020	0.000	1.734	2.439	1.0	2.0
Geo Region	Isolated	0.083	0.401	0.070	0.150	1.545	2.185	1.0	1.0
Geo Region	North	0.316	0.521	0.111	0.183	2.709	3.973	2.0	3.0
Geo Region	South	0.179	0.472	0.006	0.010	1.717	2.272	1.0	2.0
University Rank	Low	0.161	0.288	0.000	0.063	2.133	3.156	1.0	2.0
University Rank	Medium	0.327	0.516	0.079	0.096	2.455	3.597	1.0	2.0
University Rank	High	0.221	0.492	0.067	0.094	2.182	3.079	2.0	2.0
Sector	SPS/07	0.214	0.542	0.039	0.107	2.083	3.144	1.0	2.0
Sector	SPS/08	0.241	0.426	0.053	0.069	2.273	3.275	1.0	2.0
Sector	SPS/09	0.446	0.511	0.201	0.184	2.636	3.929	1.0	2.0
Sector	SPS/10	0.343	0.546	0.192	0.032	2.647	2.939	2.0	2.0
Sector	SPS/11	0.278	0.589	0.149	0.189	2.636	2.480	2.0	2.0
Sector	SPS/12	0.070	0.139	0.039	0.000	2.176	2.842	2.0	2.0
Department	Design	0.341	0.332	0.128	0.144	1.750	3.714	1.0	2.0
Department	Economics	0.199	0.349	0.039	0.000	3.000	3.652	1.0	2.0
Department	Engineering	0.042	0.447	0.000	0.447	1.000	2.000	1.0	2.0
Department	Humanities	0.183	0.415	0.042	0.062	2.365	3.326	1.0	2.0
Department	Medicine	0.058	0.038	0.000	0.024	1.750	3.000	1.5	2.5
Department	Other	0.006	0.184	0.000	0.150	1.167	2.250	1.0	2.0
Department	Political Sciences	0.243	0.476	0.070	0.000	2.184	2.554	1.0	2.0
Department	Psychology	0.498	0.497	0.180	0.201	1.750	3.600	1.5	2.0
Department	Social Sciences	0.318	0.490	0.101	0.163	1.780	3.028	1.0	2.0
Department	Sociology	0.357	0.716	0.126	0.266	3.422	4.528	3.0	3.0
Level	Assistant professor	0.136	0.600	0.035	0.092	1.710	2.709	1.0	2.0
Level	Associate professor	0.308	0.436	0.107	0.088	2.577	3.480	1.0	2.0
Level	Full professor	0.346	0.337	0.091	0.094	2.618	4.035	2.0	3.0
Level change	Same_level	0.272	0.435	0.053	0.062	2.281	3.164	1.0	2.0
Level change	Not_same_level	0.241	0.564	0.077	0.150	2.296	3.394	1.0	2.0

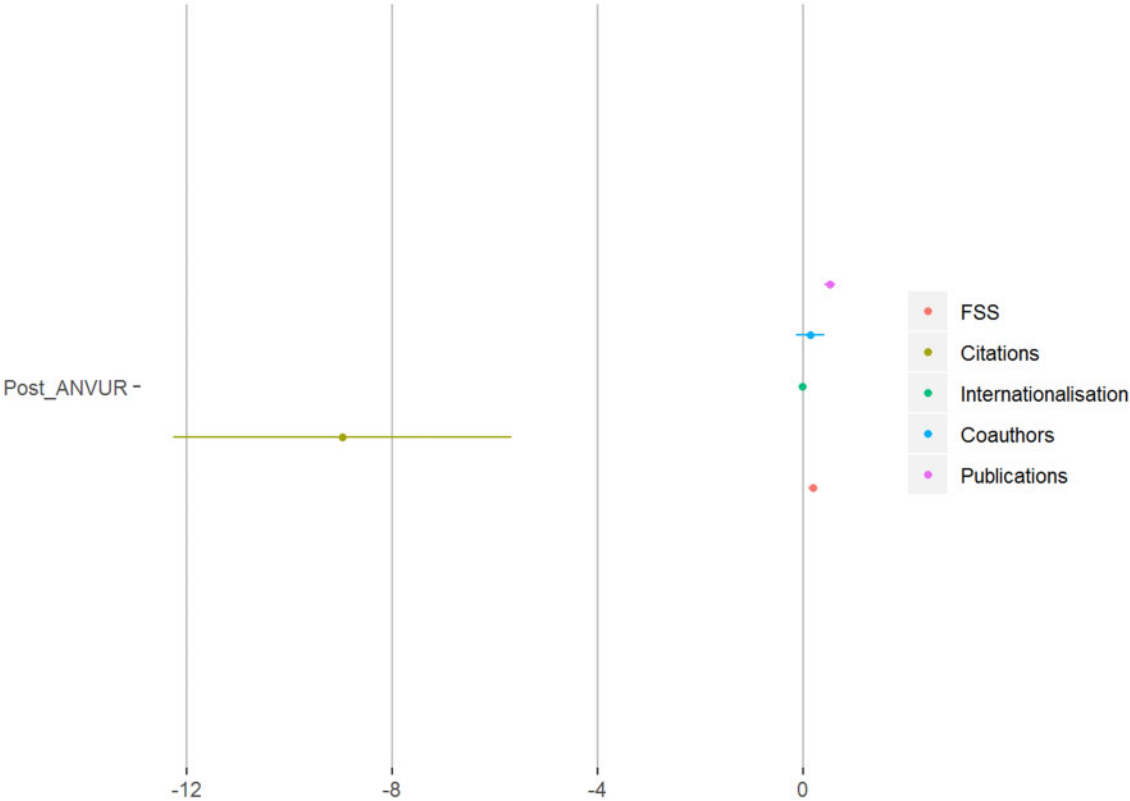


Figure 2. Baseline repeated measures mixed effect analysis to check overall ANVUR effect on research productivity (different dependent variables presented).

Table 2. Comparative table of repeated measures mixed effect analysis to check overall ANVUR effect on research productivity (dependent variable FSS).

	Gender	Sector	Level	Level 2010	Level Change	Uni Rank	Geo-region	Department	Full model
Constant	0.23 (0.09)**	0.28 (0.09)**	0.29 (0.09)**	0.20 (0.09)*	0.32 (0.07)***	0.08 (0.23)	0.28 (0.12)*	0.38 (0.48)	0.05 (0.59)
Gender M	0.09 (0.10)								0.06 (0.11)
Post-ANVUR	0.14 (0.08)	0.22 (0.08)**	0.28 (0.08)***	0.25 (0.07)***	0.11 (0.06)	0.20 (0.19)	0.12 (0.10)	-0.09 (0.40)	0.04 (0.49)
Gender M × post-ANVUR	0.07 (0.10)								0.10 (0.11)
SPS/08		-0.04 (0.12)							-0.02 (0.13)
SPS/09		0.16 (0.15)							0.20 (0.16)
SPS/10		-0.01 (0.23)							-0.02 (0.24)
SPS/11		0.12 (0.27)							0.12 (0.29)
SPS/12		-0.25 (0.24)							-0.24 (0.25)
SPS/08 × post-ANVUR		-0.03 (0.11)							-0.03 (0.12)
SPS/09 × post-ANVUR		-0.16 (0.14)							-0.20 (0.15)
SPS/10 × post-ANVUR		0.05 (0.21)							0.01 (0.23)
SPS/11 × post-ANVUR		-0.08 (0.24)							-0.06 (0.26)
SPS/12 × post-ANVUR		-0.11 (0.23)							-0.05 (0.24)
Associate professors			-0.07 (0.12)						0.06 (0.25)
Full professors			-0.03 (0.13)						0.15 (0.47)
Associate prof. × post-ANVUR			-0.08 (0.11)						-0.33 (0.24)
Full prof. × post-ANVUR			-0.23 (0.12)*						-0.74 (0.44)
Associate prof. 2010				0.11 (0.13)					0.02 (0.26)
Full prof. 2010				0.07 (0.14)					-0.17 (0.49)
Unknown 2010				0.13 (0.15)					0.30 (0.26)
Associate prof. 2010 × post-ANVUR				-0.14 (0.12)					0.20 (0.25)
Full prof. 2010 × post-ANVUR				-0.22 (0.13)					0.52 (0.46)
Unknown 2010 × post-ANVUR				0.11 (0.14)					-0.22 (0.24)
Level changed from 2010					-0.08 (0.10)				-0.28 (0.27)
Level changed × post-ANVUR					0.17 (0.09)				0.43 (0.26)
Medium rank uni						0.22 (0.25)			0.23 (0.27)
High rank uni						0.21 (0.24)			0.18 (0.28)
Medium rank uni × post-ANVUR						-0.05 (0.21)			-0.15 (0.23)
High rank uni × post-ANVUR						0.01 (0.20)			-0.06 (0.24)
Isolated							-0.10 (0.30)		-0.25 (0.35)
North							-0.00 (0.15)		-0.05 (0.17)
South							0.05 (0.19)		0.07 (0.20)
Isolated × post-ANVUR							0.07 (0.28)		0.20 (0.32)
North × post-ANVUR							0.10 (0.12)		0.10 (0.14)
South × post-ANVUR							0.05 (0.16)		0.04 (0.18)
Economics								-0.43 (0.54)	-0.40 (0.58)
Engineering								-0.24 (0.69)	-0.10 (0.72)
Humanities								-0.12 (0.49)	0.06 (0.53)
Medicine								-0.35 (0.66)	-0.17 (0.70)
Other								-0.37 (0.61)	-0.32 (0.66)
Political Sci.								-0.03 (0.50)	0.08 (0.54)
Psychology								0.21 (0.67)	0.47 (0.75)
Social Sci.								-0.03 (0.50)	0.08 (0.53)

(continued)

Table 2. (continued)

	Gender	Sector	Level	Level 2010	Level Change	Uni Rank	Geo-region	Department	Full model
Sociology									
Economics × post-ANVUR								0.05 (0.52)	0.20 (0.56)
Engineering × post-ANVUR								0.50 (0.46)	0.43 (0.50)
Humanities × post-ANVUR								0.31 (0.74)	0.12 (0.78)
Medicine × post-ANVUR								0.25 (0.41)	0.11 (0.44)
Other × post-ANVUR								0.13 (0.60)	-0.06 (0.63)
Political Sci. × post-ANVUR								0.28 (0.52)	0.21 (0.57)
Psychology × post-ANVUR								0.26 (0.41)	0.20 (0.45)
Social Sci. × post-ANVUR								-0.11 (0.61)	-0.43 (0.67)
Sociology × post-ANVUR								0.26 (0.41)	0.14 (0.44)
AIC	2,158.76	2,185.20	2,163.24	2,168.56	2,158.26	2,166.65	2,173.92	2,178.57	2,284.67
BIC	2,191.64	2,255.65	2,205.51	2,220.23	2,191.14	2,208.92	2,225.59	2,286.43	2,551.98
Log Likelihood	-1,072.38	-1,077.60	-1,072.62	-1,073.28	-1,072.13	-1,074.32	-1,075.96	-1,066.28	-1,085.33
Num. obs.	810	810	810	810	810	810	810	804	804
Num. groups: id: uniDept	592	592	592	592	592	592	592	587	587
Num. groups: uniDept	116	116	116	116	116	116	116	114	114
Var: id: uniDept (Intercept)	0.68	0.68	0.66	0.66	0.68	0.68	0.68	0.68	0.66
Var: uniDept (Intercept)	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.08
Var: Residual	0.28	0.28	0.28	0.28	0.27	0.28	0.28	0.28	0.29

***p > 0.001, **p > 0.01, *p > 0.05.

publications (Fig. 2). Results showed that the FSS increased significantly after ANVUR. We found a lower rate of citations after ANVUR, but this simply reflects the dependence of citations on time (Wang 2013).

Unlike previous studies, which suggested that the internationalization of scientific collaborations has recently increased (Akbaritabar et al. 2018; Leydesdorff et al. 2014), we found that the internationalization of this most productive subset of Italian sociologists actually decreased after ANVUR. However, the number of co-authors had increased. Note that the number of co-authorships could increase due to various factors, including the increasing appeal of team science also in the social sciences (Greene 2007; Leahey 2016). The number of publications by Italian sociologists, which we modelled with a *negative binomial distribution* to accommodate for the count nature of data, increased after ANVUR. Note that these were baseline (NULL) models with the same random structure as those described in the Methods section, but did not include any fixed effects. This was to explore general trends after ANVUR (see Supplementary Appendix Table A.5 for further detail on these models).

We then considered the effect of ‘within’ and ‘between’ groups on FSS and the *number of papers*, including the academic level and its change from 2010 to 2016. We also considered each subject’s scientific disciplinary sector, gender, academic level in 2010 and in 2016, promotion, university ranks, geographical regions, and departments, which were indicated in the *column* names in the tables of results. Differing categories of each variable are indicated in the *row* names.

Table 2 shows our results with FSS as a dependent variable.¹⁰ In general, the FSS of sociologists increased after ANVUR for most models while being statistically significant only for *sector*, *level* and *level in 2010* (see third row, i.e. *Post-ANVUR*). When considering *within-group* effects (see × *post-ANVUR* rows), *full professors* had a significantly decreasing FSS after ANVUR only for academic level.¹¹

We then considered the *total number of publications* as a dependent variable rather than the FSS presented in Table 3. While in all models (except the model evaluating departments and the full model) the total number of publications significantly increased after ANVUR (see third row, i.e. *Post-ANVUR*), we did not find any specific *within* or *between* group effect.¹²

Given that the *total number of publications* has a *count* nature and to control for the robustness of our results, we ran *negative binomial models* (see Supplementary Appendix Table B.6). For sociologists from Northern universities and who were full professors in 2010, the number of publications was significantly higher than those from universities located in central regions and sociologists who were assistant professor in 2010, respectively. In the within-group comparison, full professors had a higher number of publications after ANVUR.

As previously mentioned, while ANVUR ranked some Italian journals and assigned categories such as *Fascia* (either Fascia A, B, or C (ANVUR 2012)), we assigned *non-Fascia* to all remaining Italian or international journals. From a total of 2,133 papers in our sample, there were 350 (16.40 per cent) *Fascia* articles, while there were 1,783 (83.60 per cent) *non-Fascia* articles. We used these categories in a separate set of logistic models to calculate the probability of publication in ‘*fascia*’ versus ‘*non-fascia*’ journals before and after ANVUR.¹³ We tried to verify whether sociologists published differently in these journals. We hypothesized that under assessment, sociologists could be induced to target more national journals given that

Table 3. Comparative table of repeated measures mixed effect analysis to check overall ANVUR effect on research productivity (dependent variable number of papers published).

	Gender	Sector	Level	Level 2010	Level Change	Uni Rank	Georegion	Department	Full model
Constant	1.47 (0.30)***	1.61 (0.31)***	1.40 (0.31)***	1.48 (0.30)***	1.75 (0.25)***	1.23 (0.80)	1.48 (0.40)***	1.18 (1.62)	-0.15 (1.89)
Gender M	0.39 (0.36)								0.28 (0.39)
Post-ANVUR	1.28 (0.31)***	1.39 (0.31)***	1.29 (0.32)***	1.16 (0.30)***	1.35 (0.25)***	1.88 (0.81)*	1.07 (0.41)**	2.46 (1.63)	1.63 (1.99)
Gender M × post-ANVUR	0.29 (0.39)								0.30 (0.43)
SPS/08		-0.00 (0.43)							0.10 (0.45)
SPS/09		0.16 (0.54)							0.09 (0.56)
SPS/10		0.32 (0.78)							0.30 (0.81)
SPS/11		0.51 (0.94)							0.75 (1.00)
SPS/12		0.35 (0.80)							0.21 (0.82)
SPS/08 × post-ANVUR		0.25 (0.46)							0.23 (0.48)
SPS/09 × post-ANVUR		0.59 (0.57)							0.50 (0.60)
SPS/10 × post-ANVUR		-0.51 (0.85)							-0.42 (0.89)
SPS/11 × post-ANVUR		-0.95 (0.98)							-0.99 (1.06)
SPS/12 × post-ANVUR		-0.62 (0.92)							-0.46 (0.96)
Associate professors			0.47 (0.43)						0.45 (0.85)
Full professors			0.64 (0.44)						0.76 (1.56)
Associate prof. × post-ANVUR			0.23 (0.45)						-0.29 (0.96)
Full prof. × post-ANVUR			0.47 (0.47)						-0.76 (1.76)
Associate prof. 2010				0.67 (0.45)					0.21 (0.89)
Full prof. 2010				0.49 (0.48)					-0.25 (1.66)
Unknown 2010				-0.04 (0.53)					0.38 (0.87)
Associate prof. 2010 × post-ANVUR				0.61 (0.49)					1.02 (0.99)
Full prof. 2010 × post-ANVUR				0.52 (0.52)					1.27 (1.86)
Unknown 2010 × post-ANVUR				0.62 (0.57)					0.01 (0.98)
Level changed from 2010					-0.15 (0.36)				-0.39 (0.92)
Level changed × post-ANVUR					0.29 (0.39)				0.63 (1.02)
Medium rank uni						0.55 (0.86)			0.45 (0.91)
High rank uni						0.40 (0.85)			0.34 (0.93)
Medium rank uni × post-ANVUR						-0.26 (0.86)			-0.04 (0.94)
High rank uni × post-ANVUR						-0.54 (0.84)			-0.86 (0.97)
Isolated									-0.16 (1.17)
North							-0.18 (1.01)		0.45 (0.54)
South							0.65 (0.48)		-0.15 (0.65)
Isolated × post-ANVUR							-0.01 (0.61)		-0.49 (1.23)
North × post-ANVUR							-0.21 (1.06)		0.77 (0.55)
South × post-ANVUR							0.74 (0.48)		-0.23 (0.69)
Economics							-0.28 (0.64)	0.85 (1.84)	1.22 (1.85)
Engineering								0.07 (2.26)	0.18 (2.29)
Humanities								0.63 (1.66)	1.02 (1.68)
Medicine								0.32 (2.24)	0.95 (2.27)
Other								-0.08 (2.07)	0.32 (2.09)
Political Sci.								0.71 (1.69)	1.03 (1.71)
Psychology								-0.08 (2.27)	0.82 (2.42)

(continued)

Table 3. (continued)

	Gender	Sector	Level	Level 2010	Level Change	Uni Rank	Georegion	Department	Full model
Social Sci.									
Sociology									
Economics × post-ANVUR								0.03 (1.67)	0.43 (1.67)
Engineering × post-ANVUR								1.65 (1.74)	1.71 (1.73)
Humanities × post-ANVUR								-0.90 (1.88)	-1.09 (1.97)
Medicine × post-ANVUR								-1.67 (2.93)	-0.08 (3.00)
Other × post-ANVUR								-1.02 (1.67)	-0.83 (1.76)
Political Sci. × post-ANVUR								-0.87 (2.40)	-0.74 (2.49)
Psychology × post-ANVUR								-1.32 (2.15)	-0.44 (2.26)
Social Sci. × post-ANVUR								-1.75 (1.69)	-1.08 (1.79)
Sociology × post-ANVUR								0.01 (2.46)	0.43 (2.64)
AIC								-0.73 (1.66)	-0.26 (1.74)
BIC								-0.84 (1.69)	-0.96 (1.78)
Log Likelihood								4,088.31	4,089.59
Num. obs.								4,196.17	4,356.90
Num. groups: id: uniDept								-2,021.15	-1,987.80
Num. groups: uniDept								804	804
Var: id: uniDept (Intercept)								592	587
Var: uniDept (Intercept)								116	114
Var: Residual								4.20	4.17
								0.61	0.29
								5.46	5.46

***p < 0.001, **p < 0.01, *p < 0.05.

these were considered influential for evaluation (i.e. being listed in the 'Fascia' journals).

Table 4 shows that Italian sociologists indeed preferably published in 'Fascia' journals after ANVUR. However, this probability increased less in cases of SPS/10 compared with SPS/07 and for sociologists working in Northern universities. This probability remained statistically significant even in our full model, after controlling for all different variables. To further control for the robustness of these results, we ran Bayesian models (Figs. 3–5).¹⁴ Our results confirmed the observed trend in the baseline models. Considering that our sample covered only journals and book series indexed in Scopus (e.g. international and national journals including 50 per cent of *Fascia* journals), it is reasonable to suppose that other 'active' sociologists targeted other journals and book series not indexed in Scopus but perhaps included in 'fascia' journals.

5. Discussion and conclusions

This article aims to provide a quantitative analysis of the potential impact of ANVUR and VQR 2004–10 research assessment on Italian sociologists' publications. We considered sociologists worth studying, as this community involves the co-existence of different sub-communities, some closer to more qualitative sciences, while others to hard sciences (Akbaritabar et al. 2020). Considering that the assessment followed *informed peer review* in a context without any prior experience or consensual evaluation standards, it was interesting to examine scientists' reactions (Abramo and D'Angelo 2011a). Indeed, besides the pros and cons of this assessment (Abramo and D'Angelo 2017; Baccini et al. 2020; Baccini and De Nicolao 2016; Franceschini and Maisano 2017), evaluating the research output of scientists in similar periods can provide a picture of the interplay between institutional pressure and endogenous patterns of behaviour. This is why, regardless of detail, the VQR assessment by ANVUR could be considered the first policy experiment on 'nudging' Italian scientists towards higher scientific productivity (Ancaiani et al. 2015; Benedetto et al. 2017a, 2017b; Bertocchi et al. 2015; Bonaccorsi 2018).

It is also worth noting that VQR assessment embodied a certain degree of institutional ambiguity (Boffo and Moscati 1998). The methods and procedures used by ANVUR for sociologists included redundant signals, sometimes even contradictory, for example, discriminating publication outlets according to their quality in principle, at the same time restricting the 'Fascia' list (top quality) and excluding many renowned international journals. This would suggest that certain endogenous properties of the Italian academic system, such as a well-established system of national and local publication outlets, possibly controlled by close colleagues (Miniati and Pezzoni 2020), may have discouraged international publications without penalizing promotion and careers (Abramo et al. 2017). Furthermore, considering that scientists have their own personal career agenda, influenced by local issues and that there are fixed costs and endogenous forces that constrain productivity, it is not surprising that institutional pressures were individually interpreted differently. This lack of coherence between performance evaluation and career advancement (Abramo et al. 2014) has led some observers even to question whether the Italian academic system is prepared to reward any 'performance based' scheme (Bertocchi et al. 2015). Although the assessment had certainly positive implications, that is, establishing a competitive resource allocation scheme for university structures, its signalling function of the importance of high-quality

Table 4. Logistic regression models with differing control variables on publications in fascia A journals.

	Baseline	Gender	Sector	Level	Level 2010	Level Change	Uni Rank	Georegion	Department	Full model
Constant	-1.23 (0.11)***	-1.33 (0.15)***	-1.15 (0.15)***	-1.20 (0.15)***	-1.16 (0.14)***	-1.30 (0.13)***	-0.95 (0.34)**	-0.77 (0.19)***	-17.69 (81.99)	-16.91 (1926.80)
Post-ANVUR	0.44 (0.13)***	0.44 (0.13)***	0.45 (0.13)***	0.45 (0.13)***	0.44 (0.13)***	0.44 (0.13)***	0.45 (0.13)***	0.45 (0.13)***	0.44 (0.13)***	0.45 (0.13)***
Gender M		0.17 (0.16)								0.23 (0.17)
SPS/08			-0.18 (0.19)							-0.15 (0.19)
SPS/09			0.05 (0.23)							0.15 (0.23)
SPS/10			-0.89 (0.43)*							-0.80 (0.44)
SPS/11			0.29 (0.36)							0.28 (0.36)
SPS/12			0.19 (0.40)							0.30 (0.38)
Associate professors				-0.14 (0.19)						-0.29 (0.44)
Full professors				0.06 (0.20)						-0.21 (0.83)
Associate prof. 2010					-0.29 (0.21)					-0.12 (0.46)
Full prof. 2010					-0.00 (0.22)					0.27 (0.86)
Unknown 2010					-0.07 (0.23)					-0.44 (0.45)
Level changed from 2010						0.17 (0.16)				0.45 (0.47)
Medium rank uni							-0.22 (0.37)			-0.09 (0.36)
High rank uni							-0.36 (0.36)			-0.11 (0.36)
Isolated								-0.52 (0.44)		-0.34 (0.42)
North								-0.70 (0.24)**		-0.72 (0.21)***
South								-0.42 (0.29)		-0.47 (0.26)
Economics									16.55 (81.99)	16.28 (1926.80)
Engineering									16.15 (81.99)	15.52 (1926.80)
Humanities									16.72 (81.99)	16.42 (1926.80)
Medicine									16.45 (81.99)	16.35 (1926.80)
Other									16.56 (81.99)	15.98 (1926.80)
Political Sci.									16.20 (81.99)	15.79 (1926.80)
Psychology									16.88 (81.99)	16.52 (1926.80)
Social Sci.									16.43 (81.99)	16.14 (1926.80)
Sociology									16.46 (81.99)	16.33 (1926.80)
AIC	1,058.38	1,059.31	1,060.88	1,061.29	1,062.30	1,059.29	1,061.23	1,055.53	1,060.42	1,068.36
BIC	1,077.76	1,083.54	1,104.50	1,090.36	1,096.22	1,083.52	1,090.31	1,089.45	1,123.32	1,213.51
Log Likelihood	-525.19	-524.66	-521.44	-524.64	-524.15	-524.64	-524.62	-520.76	-517.21	-504.18
Num. obs.	940	940	940	940	940	940	940	940	933	933
Num. groups: id: uniDept	592	592	592	592	592	592	592	592	587	587
Num. groups: uniDept	116	116	116	116	116	116	116	116	114	114
Var: id: uniDept (Intercept)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Var: uniDept (Intercept)	0.23	0.22	0.18	0.22	0.23	0.22	0.19	0.18	0.11	0.00

***P < 0.001, **P < 0.01, *P < 0.05.

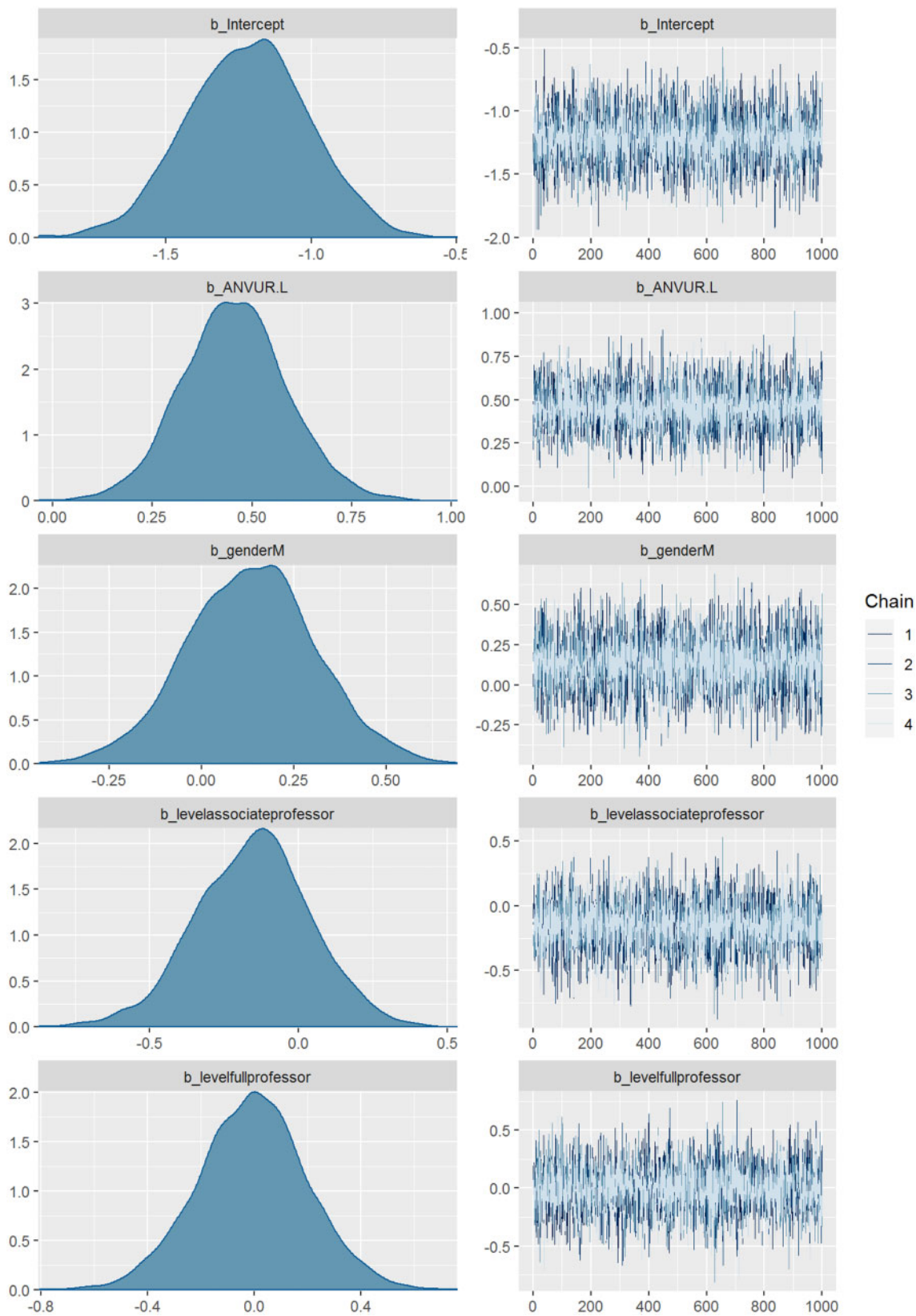


Figure 3. Bayesian model of publications in fascia journals controlling the robustness of previous models.

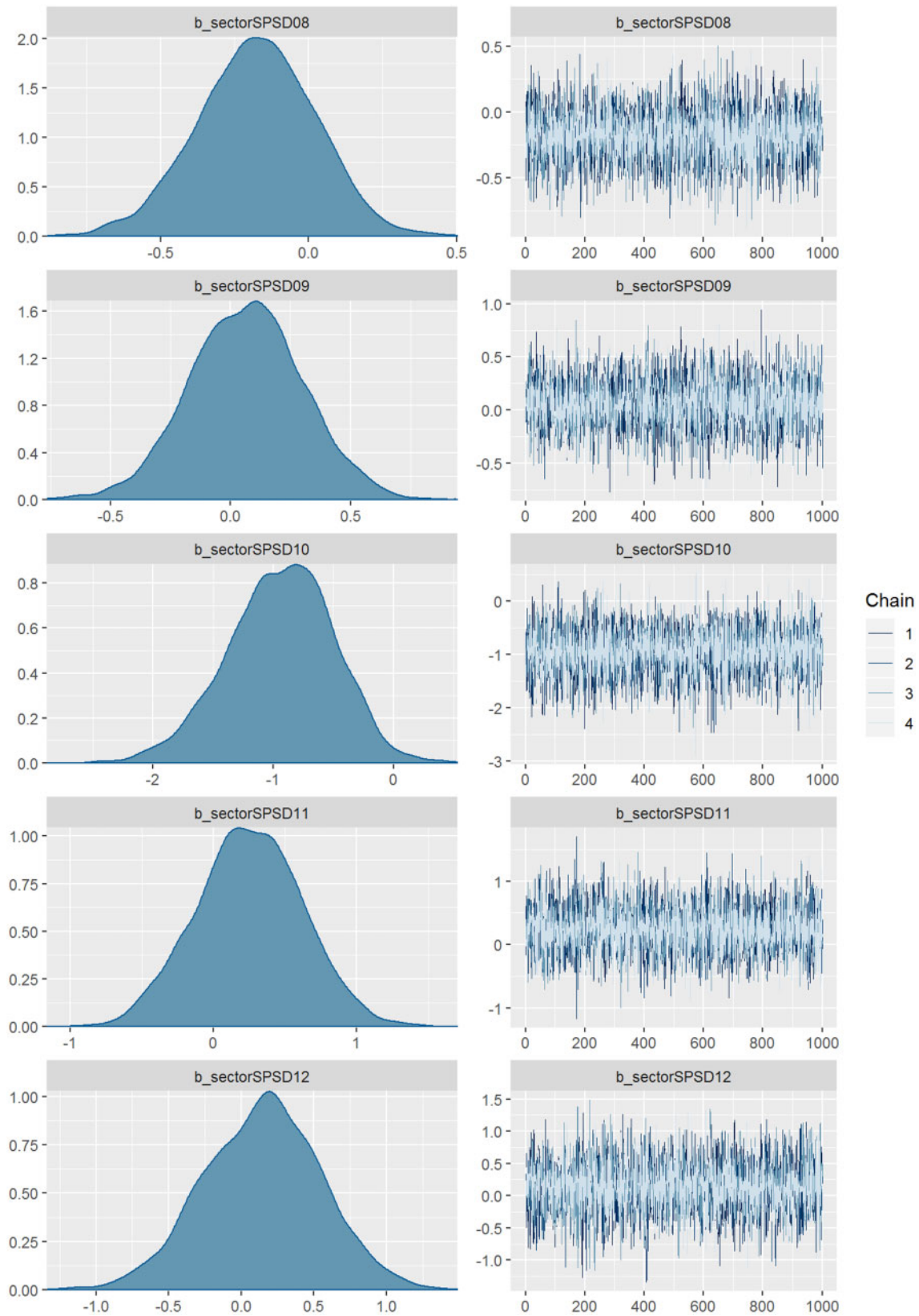


Figure 4. Bayesian model of publications in fascia journals controlling the robustness of previous models.

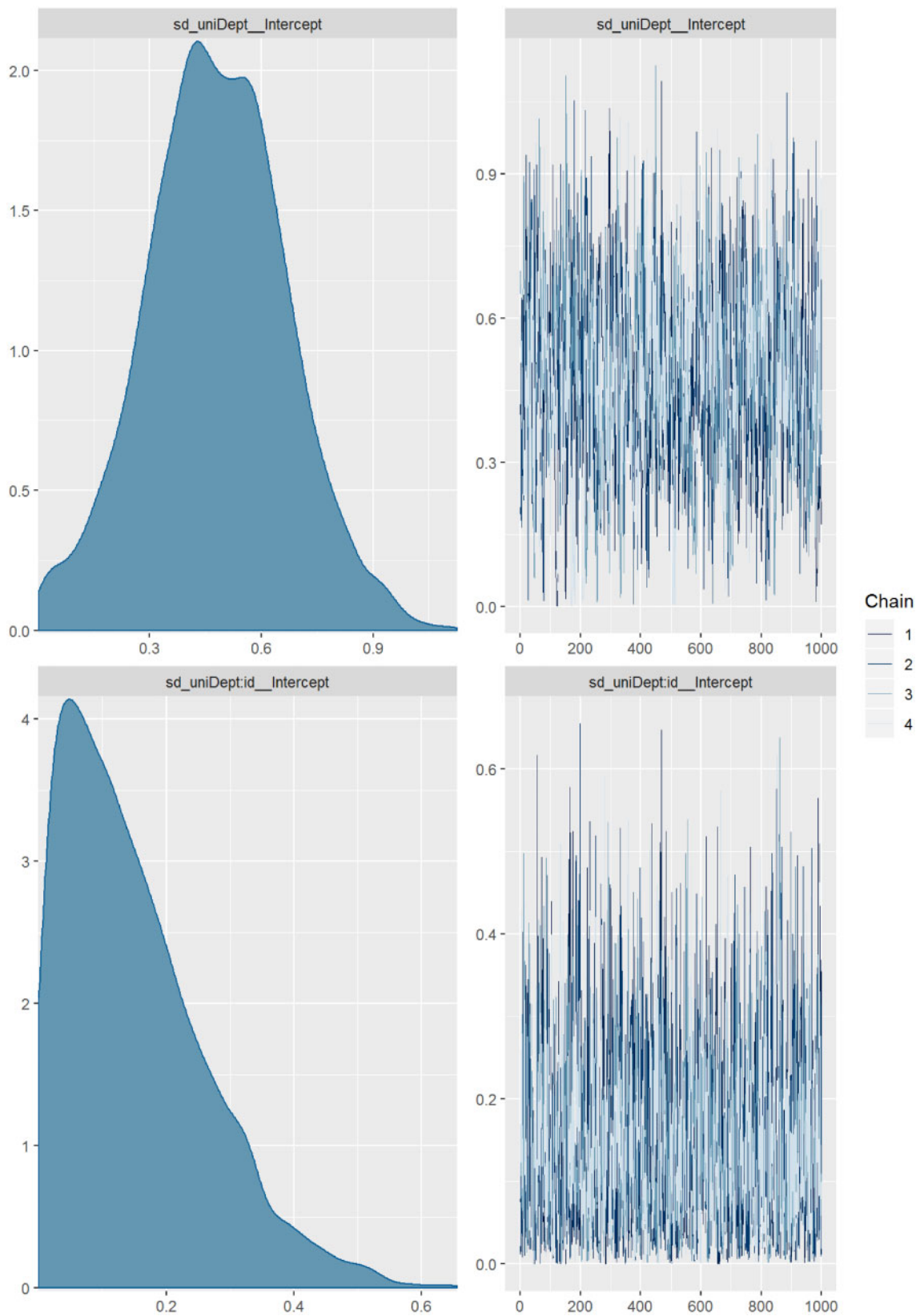


Figure 5. Bayesian model of publications in fascia journals controlling the robustness of previous models.

research was limited, with adaptive responses that probably reflected the rigidities and constraints of short-term adjustments.

Our findings suggest that ANVUR and specifically the VQR 2004–10 research assessment possibly stimulated strategic behaviour among Italian sociologists towards publishing more in ‘*Fascia*’ journals. Observed trends were consistent even after controlling for multiple contextual variables and analytical techniques. This is in line with what [Rijke et al. \(2016\)](#) hypothesized as *strategic* behaviour, where researchers might try to use metrics and quantitative evaluation procedures as a signal to target outlets, which are considered more influential in evaluation. It is probable that the ambiguity of institutional signals of the quality of publication outlets, the artificial definition of ‘*Fascia*’ journals, often regardless of the ‘objective’ prestige of outlets, and the mismatch between VQR 2004–10 and overlapping national qualification procedures were used strategically by academics ([Jonkers and Zacharewicz 2016](#)).

Secondly, while attempts at using VQR 2004–10 to link research assessment to resource allocation at the university level have been successfully pursued by MIUR, positive and negative rewards have not affected individual academics. In Italy, universities could exploit their relative degree of autonomy on careers and promotion regardless of any performance indicator of applicants. Thus, comparative studies on similar EU research ([Sandström and Van den Besselaar 2018](#)), in which universities are embedded in different institutional regimes (e.g. more or less autonomy and centralized/decentralized university systems), could improve our understanding of the complex interplay between institutional policies and individual behaviour ([Provati et al. 2012](#)).

It is therefore worth noting that initiatives aligning assessment and promotions at different academic organizational levels and based on more transparent evaluation procedures could help reduce this institutional ambiguity. Systematic research comparing trajectories and adaptations in different fields could be used to understand the interplay of pre-existing endogenous forces and the different malleability of specific communities ([Akbaritabar et al. 2020](#)). This would help to contextualize evaluation methods (e.g. bibliometrics, informed peer review and the time scale of evaluations) at least on context-specific ([Zacharewicz et al. 2019](#)), if not on individual characteristics (see the case of gender in [Marini and Meschitti \(2018\)](#), see also [Abramo et al. \(2014\)](#), (2015); [Jonkers and Zacharewicz \(2016\)](#)).

Finally, certain limitations in this study should be considered. Firstly, while Scopus covers some international and national outlets, including international book series, most Italian sociologists consider local outlets not indexed in Scopus and so excluded from our sample ([Bertocchi et al. 2015](#)). By including datasets covering national publications in more detail (e.g. Google Scholar or MIUR database), future studies could verify whether Italian sociologists reacted to the institutional pressure of ANVUR by increasing publications in these outlets. Unfortunately, as far as we know, MIUR has not made their evaluation data publicly available.

Furthermore, it is worth mentioning that Scopus is increasing its step-by-step coverage with implications on the number and type of products. This in turn affects the observed level of scientist production as reflected in our and similar samples. Controlling for these changes is impossible without proper access to micro and retrospective data on initial journal coverage. Secondly, our analysis of the effects of competition for promotion and career on publication patterns is incomplete. We were only able to reconstruct academics who were eventually promoted during the assessment ([Abramo et al. 2014](#)), while we did not have any data on candidates who applied

for the same positions but were unsuccessful. This means that the effect of competition for promotion and career may be over-estimated in our analysis.

In addition, other exogenous factors might have affected scientist behaviour, which we could not control. One of these factors was probably the introduction of the national scientific qualification as a requirement for promotion and careers ([Abramo and D’Angelo 2015](#); [Marzolla 2016](#)), which overlapped with the establishment of ANVUR and the research assessment. Furthermore, establishing exactly when ANVUR and VQR 2004–10 have started to affect academics and estimating their reaction in a common time frame is difficult ([Hicks 2017](#)). For instance, the tax deduction incentives that Italian government introduced to encourage academics—and other highly skilled nationals—residing abroad to return to the country could also have affected the rate of co-authorship observed in our study. Given that simultaneous changes in multiple policies and their often complex effects on adaptive behaviour ([Aagaard and Schneider 2017](#); [Gläser 2017](#)), our results should be interpreted cautiously and any causal relation based on our correlational research design should be avoided.

Finally, we used a *nested membership random effects structure* which enabled studying organizational contexts along with other fixed effects, while ruling out individual subjects’ random noise (see [Baayen et al. \(2008\)](#) for a discussion). However, there are other methodological possibilities to study the effect of policy changes in scientist’s behaviour, including *Dif in Dif and Regression Discontinuity Design* ([Seeber et al. 2019](#)). These efforts might help provide more insights into the effect of ANVUR on scientist reactions.

Acknowledgements

We would like to thank Marco Seeber, Peter van den Besselaar and Benedetto Lepori for their comments on an earlier version of this manuscript. We would also like to thank the editor and two anonymous reviewers for their suggestions.

Supplementary data

Supplementary data is available at *Science and Public Policy Journal* online.

Conflict of interest statement. None declared.

Code and data availability

The analysis replication script is publicly available under DOI: <https://doi.org/10.5281/zenodo.3974108>.

Notes

1. Note that the previous *Valutazione Triennale della Ricerca* (Triennial research evaluation) carried out in 2006 for the period 2001–3 had limited practical consequences in comparison to the UK RAE, which originally inspired it, and to VQR 2004–10 ([Jonkers and Zacharewicz 2016](#)).
2. Science, technology, engineering, and mathematics.
3. As an example of this pressure, we can point to debates on Italian publications, blogs and media such as: [Cassese \(2013\)](#), [Sociologia \(2014\)](#), and [AIS \(2014\)](#)
4. These are the scientific disciplinary sectors established by MIUR: General sociology (SPS/07), Sociology of culture and communication (SPS/08), Economic sociology (SPS/09), Environmental sociology (SPS/10), Political sociology (SPS/11), and Sociology of law and social change (SPS/12).

- , and De Nicolao, G. (2016) 'Do They agree? Bibliometric Evaluation versus Informed Peer Review in the Italian Research Assessment Exercise', *Scientometrics*, 108/3: 1651–71.
- , Nicolao, G. D., and Petrovich, E. (2019) 'Citation Gaming induced by Bibliometric Evaluation: A Country-Level Comparative Analysis', *Plos One*, 14/9: e0221212.
- Benedetto, S., Checchi, D., Graziosi, A., et al. (2017a) 'Comments on the Correspondence "On Tit for Tat: Franceschini and Maisano versus ANVUR regarding the Italian Research Assessment Exercise VQR 2011-2014"', *J. Informetr.*, 11 (2017), 783-787', *Journal of Informetrics*, 11: 838–40.
- , ———, ———, et al. (2017b) 'Comments on the Paper "Critical Remarks on the Italian Assessment Exercise"', *Journal of Informetrics*, 11 (2017) and pp. 337–357', *Journal of Informetrics*, 11/2: 622–4.
- Bertocchi, G., Gambardella, A., Jappelli, T., et al. (2015) 'Bibliometric Evaluation vs. Informed Peer Review: Evidence from Italy', *Research Policy*, 44/2: 451–66.
- Boffo, S., and Moscati, R. (1998) 'Evaluation in the Italian Higher Education System: Many Tribes, Many Territories. Many godfathers', *European Journal of Education*, 33/3: 349–60.
- Bonaccorsi, A. (2018) 'Towards an Epistemic Approach to Evaluation in SSH'. In: A., Bonaccorsi (ed.) *The Evaluation of Research in Social Sciences and Humanities: Lessons from the Italian Experience*, pp. 1–32. Berlin/Heidelberg: Springer.
- Butler, L. (2003) 'Explaining Australia's Increased Share of ISI Publications—the Effects of a Funding Formula based on Publication Counts', *Research Policy*, 32/1: 143–55.
- Cassese, S. (2013) 'L'ANVUR ha Ucciso la Valutazione, Viva la Valutazione!', *Il Mulino*, 62/1: 73–9.
- Coile, R. C. (1977) 'Lotka's Frequency Distribution of Scientific Productivity', *Journal of the American Society for Information Science*, 28/6: 366–70.
- De Stefano, D., Fuccella, V., Vitale, M. P., et al. (2013) 'The Use of Different Data Sources in the Analysis of Co-authorship Networks and Scientific Performance', *Social Networks*, 35/3: 370–81.
- Dupps, W. J., and Randleman, B. J. (2012) 'The Perils of the Least Publishable Unit', *Journal of Cataract and Refractive Surgery*, 38/9: 1517–8.
- Edwards, M. A., and Roy, S. (2017) 'Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition', *Environmental Engineering Science*, 34/1: 51–61.
- Engels, T. C. E., and Guns, R. (2018) 'The Flemish Performance-based Research Funding System: A Unique Variant of the Norwegian Model', *Journal of Data and Information Science*, 3/4: 45–60.
- Faraway, J. L. (2005) *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Boca Raton: CRC Press.
- Franceschini, F., and Maisano, D. (2017) 'Critical Remarks on the Italian Research Assessment Exercise VQR 2011–2014', *Journal of Informetrics*, 11/2: 337–57.
- Geuna, A., and Piolatto, M. (2016) 'Research Assessment in the UK and Italy: Costly and Difficult, but probably Worth it (at least for a while)', *Research Policy*, 45/1: 260–71.
- Gläser, J. (2017) 'A Fight on Epistemological Quicksand: Comment on the Dispute between van den Besselaar et al. and Butler', *Journal of Informetrics*, 11/3: 927–32.
- Greene, M. (2007) 'The Demise of the Lone Author', *Nature*, 450/7173: 1165.
- Hicks, D. (2012) 'Performance-Based University Research Funding Systems', *Research Policy*, 41/2: 251–61.
- (2017) 'What Year? Difficulties in identifying the Effect of Policy on University Output', *Journal of Informetrics*, 11/3: 933–6.
- , Wouters, P., Waltman, L., et al. (2015) 'The Leiden Manifesto for Research Metrics', *Nature*, 520/7548: 429.
- Jappelli, T., Nappi, C. A., and Torrini, R. (2017) 'Gender Effects in Research Evaluation', *Research Policy*, 46/5: 911–24.
- Jonkers, K., and Zacharewicz, T. (2016) *Research Performance Based Funding Systems: A Comparative Assessment*. Luxembourg: Publications Office of the European Union, EUR 27837. EN.
- Katz, J. S., and Martin, B. R. (1997) 'What is Research Collaboration?', *Research Policy*, 26/1: 1–18.
- Leahey, E. (2016) 'From Sole Investigator to Team Scientist: Trends in the Practice and Study of Research Collaboration', *Annual Review of Sociology*, 42/1: 81–100.
- Leydesdorff, L., Park, H. W., and Wagner, C. (2014) 'International Coauthorship Relations in the Social Sciences Citation Index: Is Internationalization leading the Network?', *Journal of the Association for Information Science and Technology*, 65/10: 2111–26.
- Marini, G., and Meschitti, V. (2018) 'The Trench Warfare of Gender Discrimination: Evidence from Academic Promotions to Full Professor in Italy', *Scientometrics*, 115/2: 989–1006.
- Marzolla, M. (2016) 'Assessing Evaluation Procedures for Individual Researchers: The Case of the Italian National Scientific Qualification', *Journal of Informetrics*, 10/2: 408–38.
- Miniati, R., and Pezzoni, M. (2020) 'Social Connections and Editorship in Economics', *Canadian Journal of Economics/Revue Canadienne d'économie*, 53/3: 1292–317.
- Nederhof, A. J. (2006) 'Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review', *Scientometrics*, 66/1: 81–100.
- Nygaard, L. P. (2015) 'Publishing and Perishing: An Academic Literacies Framework for investigating Research Productivity', *Studies in Higher Education*, 1–14.
- Provati, G., Squazzoni, F., and Tosio, B. (2012) 'Did They sell Their Soul to the Devil? Some Comparative Case-Studies on Academic Entrepreneurs in the Life Sciences in Europe', *Higher Education*, 64/6: 805–29.
- Ramsden, P. (1994) 'Describing and Explaining Research Productivity', *Higher Education*, 28/2: 207–26.
- Rijkse, S. D., Wouters, P. F., Rushforth, A. D., et al. (2016) 'Evaluation Practices and Effects of Indicator Use—A Literature Review', *Research Evaluation*, 25/2: 161–169.
- Sauder, M., and Espeland, W. N. (2009) 'The Discipline of Rankings: Tight Coupling and Organizational Change', *American Sociological Review*, 74/1: 63–82.
- Sandström, U., and Van den Besselaar, P. (2018) 'Funding, Evaluation, and the Performance of National Research Systems', *Journal of Informetrics*, 12/1: 365–84.
- Scopus (2017) *Scopus Content Coverage Guide*. <https://www.elsevier.com/_data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf> accessed 7 Dec 2017.
- Seeber, M., Cattaneo, M., Meoli, M., et al. (2019) 'Self-Citations as Strategic Response to the Use of Metrics for Career Decisions', *Research Policy*, 48/2: 478–91.
- Sivertsen, G. (2016) 'Publication-Based Funding: The Norwegian Model'. In: M. Ochsner, S.E. Hug, and H.-D. Daniel, (eds) *Research Assessment in the Humanities: Towards Criteria and Procedures*, pp. 79–90. Cham: Springer.
- Snijders, T., and Bosker, R. (1999) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.
- Sociologia, P. I. (2014) *Per la sociologia (ancora sull'ASN)*. <<http://perlasociologia.blogspot.com/>> accessed 4 Sept 2018.
- Šupak Smolčić, V. (2013) 'Salami Publication: Definitions and Examples', *Biochemia Medica*, 237–241.
- Traag, V. A., Malgarini, M., and Sarlo, S. (2020) Metrics and Peer Review Agreement at the Institutional Level. ArXiv:2006.14830 [Cs]. <<http://arxiv.org/abs/2006.14830>> accessed 4 Nov 2020.
- Turri, M. (2014) 'The New Italian Agency for the Evaluation of the University System (ANVUR): A Need for Governance or Legitimacy?', *Quality in Higher Education*, 20/1: 64–82.
- van den Besselaar, P., Heyman, U., and Sandström, U. (2017) 'Perverse Effects of Output-Based Research Funding? Butler's Australian Case Revisited', *Journal of Informetrics*, 11/3: 905–18.
- Waltman, L. (2018) *Responsible Metrics: One Size doesn't fit all*. <<https://www.cwts.nl/blog?article=n-r2s294&title=responsible-metrics-one-size-doesnt-fit-all>> accessed 20 Apr 2018.
- Wang, J. (2013) 'Citation Time Window Choice for Research Impact Evaluation', *Scientometrics*, 94/3: 851–872.
- Whitley, R. (2007) 'Changing Governance of the Public Sciences'. In: R., Whitley, and J., Gläser (eds.) *The Changing Governance of the Sciences*, pp. 3–27. Dordrecht: Springer.

Zacharewicz, T., Lepori, B., Reale, E., et al. (2019) 'Performance-Based Research Funding in EU Member States—A Comparative Assessment', *Science and Public Policy*, 46/1: 105–115.

Zuur, A., Leno, E., Walker, N., et al. (2009) *Mixed Effects Models and Extensions in Ecology with R*. M., Gail, K, Krickeberg, J.M., Samet, et al. (eds.) New York, NY: Spring Science and Business Media.