

Self-citations as strategic response to the use of metrics for career decisions

Marco Seeber, Mattia Cattaneo, Michele Meoli, Paolo Malighetti

Abstract

The use of metrics for research evaluation has been related to the diffusion of questionable practices among scientists. However, there is limited knowledge on the extent to which researchers may strategically respond to metrics by adopting questionable practices, and what individual and contextual factors affect their likelihood. This article aims to fill these gaps, by studying the opportunistic use of self-citations, i.e. citations to each own work, to boost metric scores. We develop hypotheses based on sociological and economic literature exploring the factors driving scientists behaviour. We test the hypotheses in the Italian Higher Education system, where the national habilitation procedure links the possibility of career advancement to the number of publications and citations received. Our sample includes 886 scientists, representing the total population of four recruitment sectors, observed along thirteen years. We find that the introduction of metrics to guide decisions on career advancement is related to a strong and significant increase in self-citations, particularly in the social science sectors, and among scientists under stronger pressure, namely researchers and those which were under performing. Our findings suggest that, while metrics are introduced to spur virtuous behaviours, if not properly designed they can rather favour the widespread diffusion of questionable practices.

1.Introduction

The practice of allocating resources and linking career advancement to research productivity has been praised for breaking the old-boy network and promoting meritocracy in the academia. Yet in recent years concern has grown over the drawbacks of the resulting pressure for publication. Studies have shown that the ‘publish or perish’ culture and the use of metrics for research evaluation can have several detrimental effects like promoting strategic game-playing and decline of shared information (Anderson et al. 2007), harming creativity (Azoulay et al. 2011; Heinze et al. 2009), spurring plagiarism (Honig and Bedi 2012), redundant publications (Jefferson 1998), fads (Van Dalen and Klammer 2005), and elitist research detached from practical and societal concerns (De Rond and Miller 2005), discouraging non-paradigmatic or a-theoretical research, and favouring ex-post hypothesizing (Miller 2007), inflating the number of submissions (Franzoni et al. 2011) and making disappear negative results (Fanelli 2012).

The traditional belief that malpractices and misbehaviours are a prerogative of researchers not sufficiently socialized into the norms of science, e.g. junior researchers, is challenged by surveys results suggesting that malpractices are rather common (Bedeian et al. 2010; Martinson et al. 2005). Some scholars argue that due to academic competition and use of performance indicators, an increasing number of researchers may be changing conceptions of what constitutes appropriate research behaviour (Martin, 2013) and engage in questionable behaviours to a point that these become embedded in the academic professional culture (Edwards and Roy, 2016).

However, despite evidence on the unintended effects of the pressure for publication, it is not yet clear to what extent researchers adopt questionable practices as a strategic response to metrics. Moreover, we have little knowledge of what individual and contextual factors affect the likelihood of such behaviours among individual researchers (de Rijcke et al. 2015), and whether there are any disciplinary variations or not.

This article addresses these gaps by studying the responses of Italian academics to the introduction of a national habilitation procedure. While so far metrics have been employed for institutional evaluation and to *inform* decisions on individual career, this procedure *directly* links the possibility to become associate professor and full professor to number of publications and citations. We study in particular the use of self-citations, i.e. citations to one’s own work, as a shortcut to boost metric scores. We focus on self-citations for two main reasons. First, for a practical reason. Timely knowledge on the undesirable effects of employing metrics to take decisions on career is necessary. However, due to the time needed to craft and public an article, many years may be needed before being able to detect changes in the research and publication strategy or the adoption of more severe questionable behaviours. Instead, self-citations can be exploited in very short time, simply adding citations to the articles in the pipeline, and an increase in self-citations can be easily detected. Therefore, self-citations represent a valuable ‘canary in a coal mine’: a fast and visible indication on the extent to which researchers respond opportunistically, and a warning of potentially more harmful unintended effects that can occur or become visible in the longer run. The second reason for studying the abuse of self-citations is that they represent one notable typology of the emerging ‘post-production misconducts’, namely misconduct aimed at enhance publications impact. While there practices do not generate false results, they nevertheless erode the credibility of the publication system (Biagioli, 2016). Moreover, inflating self-citations is also at odd with the norm of disinterestedness that is central to the ethos of science (Merton, 1973). As such, opportunistic self-citations can signal that the norms of the scientific community are not effectively guarded, which is likely conducive to more

infringements and more serious ones, in line with the so-called 'broken window' effect (Wilson and Kelling, 1982; Keizer et al. 2008).

We build on sociological and economic studies to develop hypotheses on the individual and contextual factors that drive scientist to increase self-citations in response to the introduction of metrics-based decisions on careers. We test the hypotheses considering the scientific production of 886 scientists in the period from 2002 to 2014. The sample includes the entire population of four 'recruitment-discipline' sectors, representative of research in the natural science, medicine and social science research.

The remainder of the article is organized as follows. In Section 2 we review the literature on the drivers of scientists behaviour, and develop hypotheses on the adoption of questionable practices, namely increasing self-citations, in response to metric based habilitation. In Section 3 we present the data and methods. The empirical analysis is presented in Section 4. In the concluding section we discuss the main findings, the theoretical and policy implication, and advance proposals to counteract some of the possible unintended effects of the use of metrics.

2.Strategic responses to metrics

2.1 Quantitative indicators for research evaluation

In recent decades, quantitative indicators have proliferated and diffused for purposes of assessing scientific output, and drive the allocation of funds, hire and promote staff (Miller et al. 2005; Van Fleet et al. 2000; Harzing 2010; Lissoni et al. 2011; de Rijke et al. 2015). The use of quantitative indicators is based on the assumption that publications are proxies of scientific progress and that the number of citations received reflects their contribution to progress and advancement of science (Moed 2005). This latter assumption derives from Merton's view of citations as acknowledging the knowledge claim of the cited source (Merton, 1988; Haustein and Lariviere, 2015), and it was backed by evidence of a positive correlation between citations rates and qualitative judgement by peers (Cole and Cole, 1973; Zuckerman, 1987).

In parallel to their diffusion, a debate emerged regarding how indicators should be properly constructed, their conceptual and empirical validity, as well as their unintended effects (MacRoberts and MacRoberts 1989; Garfield 2003; Costas and Bordons 2007).

In building quantitative indicators, crucial decisions regard the type of publications to be included - e.g. articles and books, their language, national or international journals – and whether self-citations should be considered or not (Waltman, 2016). Research has shown that self-citations have a negligible effect at system level evaluation (Aksnes, 2003; Glänzel and Thijs, 2004). Instead, most scholars support that self-citations should be excluded for institutional and individual evaluation, but it is debated whether their impact is negligible (Hirsch, 2005; Engqvist and Frommen 2008) or not (Schreiber, 2007; Gianoli and Molina-Montenegro 2009). Fowler and Aksnes (2007) found that beyond a direct benefit on citations, self-citations also have an advertising effect that they estimate in additional 3.65 citations per each self-citation, and suggest to discount the citations count for this effect as well.

Some studies have explored the validity of the assumptions of quantitative indicators. On the one hand, the correlation between a journal impact factor¹ and the article's impacts is actually weak (Lozano et al. 2012) as well as the relationship between article impact and quality (Saha S. 2003;

¹ i.e. the average number of citations received per year by the journal' articles in two (or five) years after the publication.

Weingart 2005). On the other hand, recent analyses confirmed that, despite most scholars believe that citation counts are not a good proxy of scientific contribution (Aksnes and Rip, 2009), when asked about their own publications, indeed the citation counts do correspond with the degree of perceived contribution (Aksnes, 2006).

A major problem, which goes beyond the academic context, is that assessment criteria tend to become the goal themselves, replacing the original goal they intended to measure (Kerr 1975). For instance, academics in the United Kingdom changed their publication behaviour depending on whether the evaluation criteria of the research assessment exercise focused on quantity or quality of publications (Moed 2008).

Intentionally or not, the use of indicators is becoming increasingly pervasive, and affect decisions at the individual level. It has been observed that criteria employed for evaluating system and institutions tend to trickle down and become gradually employed at lower of analysis (Hammarfelt et al. 2016), while indicators for the evaluation of individual research has multiplied themselves, Wildgaard et al. (2014) recently identified 108 of them. Still, thus far decisions on individual career have been mostly *informed* by metrics, while their use to *formally guide* decisions on careers is rather recent, rare, and its effects barely unknown.

2.2 The increase of self-citations as response to metric-based evaluation

Classical sociological accounts depicted scientists' behaviour as driven by the enjoyment derived from solving 'puzzles' as well as by the recognition from peers' for achieving a discovery (Eiduson, 1962; Hagstrom; 1965). The scientists' social context was conceived as dominated by an ethos of science - characterized by prescriptions of communalism, universalism, disinterestedness and organized scepticism (Merton, 1973; Hagstrom 1974; Zuckerman, 1977). Overall, scientists' motivations and the normative environment contributed to the well-functioning of the scientific enterprise. Post-world war II research policies in western countries developed under similar beliefs that scientists, left free to pursue their curiosity, would have naturally provided the knowledge that the nation needed (Bush, 1945).

While sharing the importance of curiosity and peer recognition, since early nineties economists argued that scientists are also interested in more mundane returns, namely money, and that their behaviour is not merely driven by ethical concerns, but they are strategic in pursuing their goals (Levin and Stephan, 1991; Stephan, 1996). In a similar period, research policies have increasingly aimed to spur efficiency and performance through competition and assessment of results. Therefore, at least implicitly, policies have also experienced a similar shift in their assumptions about what drives scientists' behaviours, which are not necessarily interested in their own job and must be therefore spurred via incentives and controls (for a similar change in public policies, see: Kaboolian 1998; Vandenabeele 2007).

Different perspectives on the nature of scientists lead to different interpretations of and expectations on the adoption of questionable behaviours. Consistently with a Mertonian view of an ethical and disinterested scientist, early studies tended to blame the individual. Offenders were described as few black sheep with peculiar psychological profiles, affected by anomie or alienated (Anderson et al. 1994; Hackett 1994). Adopting a less normative stance, later studies have paid attention to more systemic, undesirable behaviours induced by an improper system of incentives (Stephan 2012). Franzoni et al. (2011), for instance, found that the number of submissions to *Science* journal increased significantly in countries that introduced monetary incentives for publications, but the rate of acceptance decreased, so that in turn the effect was only to overload

the reviewing process. Practices that are counterproductive or indifferent in terms of achieving priority in a discovery are often adopted because valuable to increase a researcher's metric scores. For instance, 'cutting' a scientific result into many (long) works is profitable in terms of number of articles and citations received instead of presenting them in a standalone work (Bornmann and Daniel 2007), homogeneity and specialization may reduce opportunity for breakthrough discoveries but are nevertheless more effective in terms of number of citations received when compared to an heterogeneous profiling (Leahey 2007), publication strategies should not be too selective in order to avoid undervaluation by bibliometric indicators (Costas and Bordons 2007).

In turn, recent evidence suggests that scientists are highly responsive to incentives. Possibly, the strategic response may include the adoption of questionable behaviours as well, like augmenting self-citations to boost metric scores. Therefore, we explore the hypothesis that:

Hypothesis 1) The introduction of metrics to guide decisions on careers increases the number of self-citations

From a strategic perspective, individuals will be particularly responsive when they have a strong need to do so. Two further hypotheses can be derived. First, since full professors have reached the highest possible rank, hence it is expected that:

Hypothesis 2) The introduction of metrics to guide decisions on careers increases the use of self-citations among scientists in the lower academic ranks (e.g. associate professors, researchers) than among full professors

Moreover, some scientists may feel already confident in their metric scores, while academics that are underperforming compared to their peers may feel the urgency to increase the citations to catch up with competitors. Thus:

Hypothesis 3) Scientists that are more cited than their peers of the same rank and recruitment sector will increase the self-citations comparatively less

Beyond the competitive pressure, the social context in which scientists are embedded may also affect scientists' strategic response. In particular, Whitley's (2007) argued that the consequences of using indicators for research evaluation depend to some extent on the social and intellectual structure of a discipline. Since social sciences and humanities display a less cohesive social structure than the natural sciences,² these disciplines are more prone to imitate research practices and norms of dominant fields, and the impact of evaluation systems will be stronger. Some studies have backed this argument. Hammarfelt's and de Rijcke's (2014) recently observed an increase in English language journal articles – the typical output of science, technology, engineering and mathematics (STEM) disciplines - among scientists in the humanities, and Aagaard et al. (2015) found that the Norwegian national indicator for research evaluation is used more frequently at individual level and is more important for judging scholarly status in the humanities. Reale and Seeber (2011) found that the introduction of a research assessment exercise in Italy had a stronger effect on the research practices of management scientists, than scientists in physics and in the biomedical science, since the latter were already accustomed to consider the number of articles in international journals and citations as proxies of scientific value. Hence, we formulate the hypothesis that:

² As a consequence of less agreement on research goals, less collaboration, large freedom for individual researchers to develop their own research agendas (Whitley, 2007)

Hypothesis 4) The introduction of metrics to guide decisions on careers increases the use of self-citations more in the social sciences than in the natural sciences and medicine

3. Data and Methods

3.1 Context of analysis

We test the hypotheses in the context of Italian higher education. This system is predominantly public – almost 95% of academics employed in public universities – and funded by public purse.

Traditionally, Italian higher education was governed by a diarchy of a centralized bureaucracy exerting procedural control through laws and regulations, and powerful chairs, which held decision making power over academic issues as well as individual careers (Giglioli 1979; Clark 1983). From mid-nineties, reforms loosely inspired by New Public Management principles aimed to change the governance of the system, from a steering “by law and regulation” to a steering “by objective” approach, where resource allocation is driven by performance and universities gain more managerial autonomy. In this context, decisions over careers moved from national to university-selected disciplinary committees. However, the devolution of powers to the universities was not accompanied by a strong enough system of accountability, and the distribution of resources have remained weakly competitive (Reale and Seeber, 2011). Possibly as a consequence, the recruitment procedures became increasingly criticized due to alleged lack of meritocracy, the high level of localism, cases of cronyism and nepotism (Morano Foadi 2006; Durante et al., 2011).

In order to address such shortcomings, in late 2010 the law 240 introduced a national habilitation procedure regulating access to professorial positions. Accordingly, committees for each disciplinary ‘recruitment sector’ can award the habilitation only if the candidate meets specific thresholds of productivity. Two different approaches were conceived. Humanities and Social Sciences disciplines adopt a so-called *non-bibliometric* approach, whereby each committee sets distinct thresholds of productivity in terms of a minimum number of publications in pre-defined ‘first class’ and ‘second class’ journals. Instead, medicine, engineering and natural science disciplines adopt a *bibliometric* approach, which prescribes that in order to obtain the habilitation as associate/full professor, the candidate should outperform the median productivity of the current body of associate/full professors in at least two out of three productivity indicators: i) number of publications, ii) number of citation received and iii) Hirsch index.³

In turn, the Italian habilitation procedure presents some unique features to explore the strategic responses to metrics and the adoption of questionable practices. First, while metrics have been traditionally used to inform decisions on individuals, instead the habilitation procedure directly affect the possibility of individual career advancement. Second, the habilitation is defined and applied at system level and to all disciplines with very similar parameters. Third, the habilitation allows to control the effects of different metric systems (see next paragraph).

3.2 Sample

³ The data on publication can be retrieved from Scopus or Web of Science databases, depending on the candidate’ choice. The individual productivity scores are normalized by taking into consideration the number of years from the first publication.

From a disciplinary perspective, the Italian HE system is organized in 367 disciplinary sectors, nested in 188 recruitment sectors, 88 macro-sectors and 14 macro-areas.⁴ We theoretically sampled four recruitment sectors of similar size, which could be representative of science' major research areas, namely natural sciences, medicine and social sciences. Moreover, we selected two sectors that have a rather similar focus of research but adopt respectively a bibliometric and a non-bibliometric approach, namely 'Economic and managerial Engineering' and 'Applied Economics'. We retrieved information on publications, citations, affiliations and career evolution of the entire population of scientists from the four sectors, for a total of 886 academics, from 2002 to 2014.⁵ Data on publications and citations have been retrieved on August 2016 from Scopus database, which is the official database employed for computing bibliometric indexes.⁶ Information on scientists' careers and affiliations were retrieved from Cineca database.⁷

Table 1 presents the main characteristics of the selected recruitment sectors.⁸

Table 1 –Characteristics of the recruitment sectors

Sector	Macro Area	habilitation	groups	academics	publications
Genetics (Sciences)	Biology	Bibliometric	42	260	5481
Economic and Managerial Engineering (Social science)	Engineering and Architecture	Bibliometric	36	221	3429
Psychiatry (Medicine)	Medicine	Bibliometric	38	258	10271
Applied Economics (Social science)	Economics and Statistics	Non-bibliometric	53	147	1384
			169	886	20565

3.3 Self-citations

As a questionable practice we consider increasing self-citations as a shortcut to boost one's own metric scores. Self-citations are defined as citations from other works co-authored by the author.⁹ The dependent variable is given by a scientist's number of self-citations in a given year i .

It is important to remark that an increase in self-citations might well indicate that the scientist's work tend to be related to its previous works more than in the past. However, a rapid increase in self-citations corresponding to the introduction of the habilitation procedure is likely indicative of a strategic use of self-citations rather than a narrowing focus of research, because some years are arguably necessary to adapt one's work.

⁴ There are: 1. Mathematics; 2. Physics; 3. Chemistry; 4. Geology; 5. Biology; 6. Medicine; 7. Agricultural Science; 8. Architecture and Civil Engineering; 9. Engineering and Informatics; 10. Humanities; 11. Philosophy, History, Pedagogy, and Psychology; 12. Law; 13. Economics and Statistics; 14. Social and Political Science.

⁵ We excluded year 2015, since at the date of the extraction information was not yet complete. For each scientist we excluded years in which the person was not affiliated to the recruitment sector (either because non tenured or affiliated to another sector) and years when the productivity was null.

⁶ The candidate can choose that the indexes are computed using Scopus either Web of Science database

⁷ Cineca is a non-profit consortium of Italian universities, national research centers, and the Ministry of Universities and Research (MIUR), which is formally in charge of collecting statistics on Italian Higher Education and research system.

⁸ Each of the four selected recruitment sectors include only one disciplinary sector.

⁹ In this respect, we adopt a narrower definition of self-citation than e.g. Aksnes (2003) and Glänzel et. al. (2004) which consider self-citation a citation from any article with at least one of the coauthor in common between the citing and the cited article.

3.4 Independent variables predicting the number of self-citations

Some 'baseline' variables are expected to predict the sheer *number* of self-citations. These include:

Recruitment sector. A dummy variable identifies the recruitment sector, in order to take into account for possibly different citation practices.

Number of products in year i. The number of self-citations is expected to be strongly related to the number of publications co-authored in a year.

Past productivity. The number of publications co-authored in the past increases the pool of articles that can be cited. Hence, a variable measures the cumulative number of publications up to year $i-1$. Moreover, we introduce a variable given by the square of the cumulative number of publications up to year $i-1$ in order to consider the expected decreasing marginal effect of an additional item to the pool of past publications.

Number of co-authors. It can be expected that the larger the number of co-authors of a publication and the lower the number of self-citations that can be included for each of them. Several reasons support this expectation: publications cannot contain too many citations, the more the co-authors and the shorter the section in which each one can self-cite her/his own work, excessive self-citation may not be appreciated by other co-authors. We construct a variable to measure the average number of co-authors for the publications of an author x in year i .

Characteristics of the individual scientist and of the context in which they work may also affect the propensity to self-cite.

Academic rank. Since self-citations can be used to increase prestige (Hutson 2006) and scientists of lower academic rank may be more in need to affirm their status, it is expected that researchers and associate professor will cite their work more than full professors. A dummy variable considers the scientist' academic rank in year i .

Citation standing. Scientists that receive many citations may be more confident on the relevance and value of their work, and be more prone to cite their own work. We introduce a variable given by the amount of citation received normalized by the citations received by the median scientists of the same rank and disciplinary group.

Local competition. Colleagues that are both in the same university, sector and rank represent the closest source of competition for prestige. Therefore, the number of local peers may relate to a stronger propensity to self-cite.

Gender. Three studies on self-citations explored gender variations and found that male scientists are more prone to self-cite than female scientists (Hutson 2006; Maliniak et al. 2013; King et al. 2016). We therefore include a dummy variable for a scientist' gender.

Professional age. The propensity to cite one' own works increases with the professional age of the author (Hutson, 2006). Possibly, this may reflect the fact that senior academics are less prone to explore new areas of research compared to their younger colleagues.

3.5 Independent variables: increase of self-citations in response to the habilitation procedure

Habilitation. We expect the introduction of habilitation criteria based on metrics to increase the propensity to self-cite (Hypothesis 1). We construct and test three different dummy variables for three potential years when the change in behaviour could start: 1) when the law was issued (2011 onwards), 2) the year after (2012), or the 3) year of the first formal habilitation procedure (2013).

Regression discontinuity design. To rule out the possibility that changes in the number of self-citations are driven by an overlapping general trend, rather than by the introduction of the habilitation criteria, we adopt a regression discontinuity design (RDD) (Imbens and Lemieux, 2008). This consists in a non-experimental setting that uses ex-post data to evaluate a program's impact in a context or a situation where units are considered treated or not, according to a certain threshold in a reference variable. This methodology is particularly suitable to study the effects of regulatory changes because it allows isolating the effects of the regulatory shift from those of a changing climate and other confounding, contemporaneous events. In our setting, the forcing variable is the entry into force of the habilitation criteria. The RDD implies the specification of models where a p-th order parametric polynomial accounts for non-linearity of the relationship between the time trend and the dependent variable. In particular, we control the counter hypothesis of a linear, quadratic and/or cubic growth throughout the period by introducing a 3-rd order polynomial centred around the year of event. The use of higher order polynomial is discouraged by recent literature (Gelman and Imbens, 2014).

Interaction between habilitation and academic rank. We test an interaction term between the habilitation and academic rank variables to explore hypothesis 2 that researchers and associate professors will increase self-citations more than full professors.

Interaction between habilitation and citation standing. We test an interaction effect between the habilitation and citation standing variables to explore hypothesis 3 that scientists that are more cited than their peers of the same rank and recruitment sector will increase the self-citations comparatively less.

Disciplinary variations. We expect that the increase of self-citations will be particularly strong in the Social Sciences (Hypothesis 4). Therefore, we introduce an interaction term between habilitation and discipline sector variables.

3.5 Analyses

The empirical analysis combines descriptive and inferential statistics.

First, we describe the longitudinal evolution in scientists' propensity to self-cite along the four recruitment sectors.

Next, we run regression models to test the factors predicting the number of self-citations and the change in self-citation behaviour related to the habilitation process. The dependent variable is given by the count (number) of self-citations of an author in a given year, and it is over dispersed (i.e. the variance increases faster than the mean). For dependent variable which is a count, negative binomial regressions are preferred to Poisson regressions for they include a distinct parameter to model over dispersion (Snijders and Bosker, 2012). Negative binomial models do not assume that observations are normally distributed around the conditional mean, therefore usual fit measures based on the assumption of normality (like R^2) cannot be used. Instead, the fit can be evaluated by indexes measuring accuracy in the predictions, e.g. of sensitivity and specificity. It is important to remind that binomial regression coefficients are exponential and multiplicative: if the coefficient for an antecedent is β , then the change in the expected number of counts for a unit change in the antecedent is given by the exponential of β .¹⁰

The data have a three-level structure, with years nested into 886 academics, nested within 169 groups. Ignoring the multilevel nature of a data structure can be perilous, e.g. leading to ecological

¹⁰ For instance: coefficient equals 0.40, exponential (0.4)=1.49, therefore the increase is 49%.

fallacy (Robinson, 2009) and inaccurate estimates (Snijders and Bosker, 2012). Instead, a multilevel regression model allows to disentangle the variance that is due to individual and contextual factors and to compute more accurate coefficients.¹¹ Therefore, we estimate the following multi-level model:

$$\text{Log}(\pi_{ijk}) = \beta_0 + \beta_1 X_{ijk} + \beta_2 X_k + u_{ijj} + u_k + e_{ijk}$$

Multicollinearity is not a major concern in our regression models because none of the variance inflation factors exceed the value of 8.17, which is below the critical cut-off of ten.

Finally, we develop a robustness tests by controlling whether the results hold when the regression models are run for each recruitment sector individually.

4. Analysis

4.1 Descriptive statistics

Table 2 displays the descriptive statistics of the dependent variable and selected independent variables.

Table 2 – Descriptive statistics. n=5,780 author_x year_i

variable		Min	Median	Mean	Max	Standard Deviation
Self-citation per year		0	2,0	6	275	14
Publications year		1	2,0	4	67	4
Cumulative production year t-1		0	21,0	37	433	47
Mean co-authors		1	5,5	6,1	22,0	3,7
Citation standing (normalized)		0	0,7	1,23	22,90	1,67
Local competitors		0	2	3	18	4
Professional age		0	16	18	59	11
rank	researchers		2.317		40%	
	associate professors		1.804		31%	
	full professors		1.659		29%	
gender	Female		1.978		34%	
	Male		3.802		66%	

The number of self-citations strongly and significantly correlates – as expected – with the number of publications (0.746**) and the cumulative past publications (0.567**), which are also highly correlated with each other (0.609**). Correlations with the other variables are also significant but of lower value (below 0.35); surprisingly, correlation is positive with the mean number of co-authors, yet correlations run within the recruitment sectors are negative or null.

Table 3 – Pearson correlations

	Self-citations per year	Publications year	Cumulative production year t-1	Mean co-authors	Habilitation 2011	In local competitors	Citation standing	Gender: Male vs Female
Self-citations per year	1	,746**	,567**	,170**	,121**	-,040**	,350**	,068**

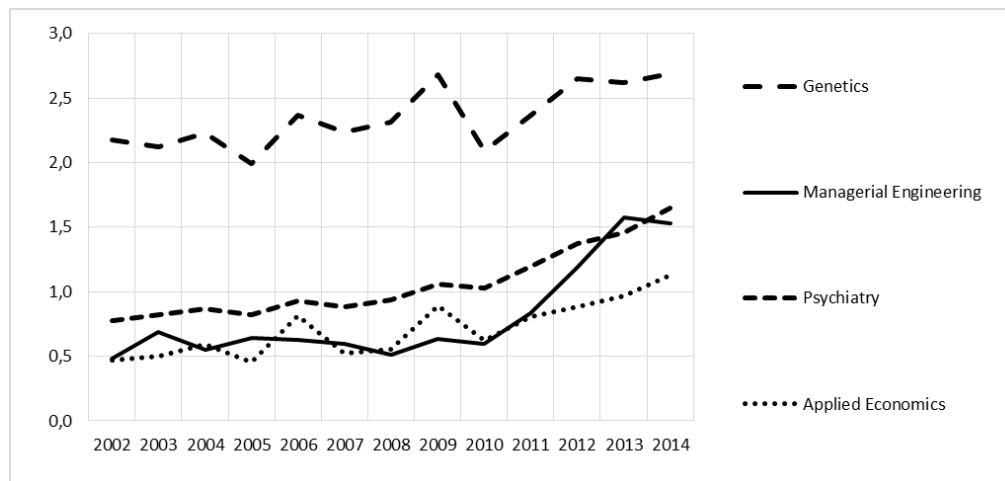
¹¹ Maximum-likelihood estimates may provide biased results for multilevel models (Snijders and Bosker, 2012). PQL estimations are employed instead, while Monte Carlo Markow Chain MCMC models are not implemented in MLwiN, the statistical software employed for the analysis.

Publications year	,746**	1	,609**	,171**	,105**	-,051**	,295**	,135**
Cumulative production year t-1	,567**	,609**	1	,270**	,067**	-,086**	,365**	,142**
Mean co-authors	,170**	,171**	,270**	1	,035**	-,049**	,031*	-,086**
Habilitation 2011	,121**	,105**	,067**	,035**	1	-0,002	-,033*	-,021
ln local competitors	-,040**	-,051**	-,086**	-,049**	-0,002	1	,038**	-,071**
Citation standing	,350**	,295**	,365**	,031*	-,033*	,038**	1	,040**
Gender: Male vs Female	,068**	,135**	,142**	-,086**	-,021	-,071**	,040**	1
ln professional age	,185**	,164**	,478**	,300**	,034*	-,029*	,130**	,033*

* Correlation is significant at the 0.05 level (2-tailed). ** Correlation is significant at the 0.01 level (2-tailed).

Figure 1 illustrates the evolution of the average number of self-citation per publication. Between 2002 and 2014 they have increased in all four sectors, although the growth appears more stable and remarkable from 2010-2011 onwards.

Figure 1 –Evolution of the average number of self-citations per publication



4.2 Regression analysis

Table 4 presents the results of four multilevel negative binomial regressions.

Model 1 includes the baseline predicting variables, and Model 2 adds the control variables. The predicting variables are significant or strongly significant, with the exception of the mean number of co-authors and gender dummy variable. Researchers and associate professors self-cite remarkably more than full professors, as well as scientists that have attracted many citations and those with more years of academic professional age. The local competitive pressure has a positive but not significant coefficients. An important remark is that the gender variable is never significant. Previous studies that found that male scientists self-cite significantly more only included a limited amount of covariates; these results suggest that if several covariates are considered then gender-related differences barely disappear.

While the variance of the dependent variable is mostly at year and author levels, it is interesting to observe that variance between groups is also significant - and not negligible - suggesting that the local norms and practices meaningfully affect self-citation practices.

Model 3 tests hypothesis 1 that self-citations increase when decisions on careers become formally affected by metrics scores. The coefficient of the variable habilitation (2011) is significant, and the

coefficient point out a +15% increase in self-citations (exponential of 0.15). The polynomial testing linear, quadratic and cubic growth are also significant – suggesting a baseline trend of growth in self-citation – which still do not absorb the habilitation effect. The effects of the other variables remain similar, with the exception of the mean number of co-authors which becomes significant, and the coefficient of the professional age, which diminishes. Alternative tests with 2012 and 2013 as habilitation years display less meaningful changes in self-citation behaviour (see appendix), suggesting that scientists have been strongly reactive in changing their self-citation practices ever since the law was issued, and that 2011 mark the turning point in the related self-citation behaviour.

The results of Model 4 partially support hypothesis 3 showing that researchers increase self-citations significantly more than full professors after the introduction of a metric based habilitation, whereas there is not a significant difference between associate and full professors. The results back hypothesis 4 that Scientists that were cited more than their peers of the same rank and recruitment sector have increased self-citations comparatively less. The effects of baseline and control variables remain similar to Model 3.

Model 5 tests whether self-citations increase more strongly for scientists in the social science than in other disciplines (hypothesis 4). For this purpose, the model includes four interaction terms between habitation and disciplines. The results show that the coefficient is the strongest for scientists in ‘economic and managerial engineering’ (+84%) - a field of research in the social science and adopting a bibliometric approach - followed by ‘applied economics’ (+29%) – a field in the social science adopting a ‘non-bibliometric’ approach, whereas the effect is not significant for the sectors in medicine and natural sciences, despite the bibliometric approach explicitly considers the number of citations as a criterion for the habilitation. In turn, the results corroborate hypothesis 4. The effects of the other variables is very similar to Model 4.

Table 4 - Regression models : number of self-citations and change in self-citation

Dependent variables	model 1: baseline variables			model 2: baseline and control variables			model 3: hp 1 - habilitation			model 4: hp 2 and h 3 : habilitation * rank & competitive position			model 5: hp 4 : disciplinary variations		
	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign
Cons	0,27	0,09	**	-1,40	0,19	***	-0,32	0,21		-0,31	0,21		0,06	0,21	
sector: Managerial Engineering	-0,64	0,13	***	-0,34	0,13	*	-0,81	0,14	***	-0,81	0,14	***	-1,24	0,14	***
sector: Psychiatry	-1,15	0,11	***	-1,07	0,11	***	-1,04	0,11	***	-1,05	0,11	***	-1,11	0,11	***
sector: Applied Economics	-1,04	0,13	***	-0,72	0,15	***	-1,25	0,15	***	-1,23	0,15	***	-1,48	0,16	***
Publications year	0,19	0,00	***	0,19	0,00	***	0,18	0,00	***	0,18	0,00	***	0,18	0,00	***
Cumulative production year t-1	0,018	0,001	***	0,012	0,001	***	0,006	0,001	***	0,007	0,001	***	0,007	0,001	***
Square: Cumul. prod. year t-1	-0,000046	0,000003	***	-0,000033	0,000004	***	-0,000023	0,000004	***	-0,000022	0,000004	***	-0,000023	0,000004	***
Mean coauthors	-0,007	0,005		-0,006	0,005		-0,014	0,005	**	-0,014	0,005	**	-0,012	0,005	*
Researcher vs Full Professor				0,30	0,08	***	-0,01	0,08		-0,06	0,09		-0,05	0,08	
Associate vs Full Professor				0,24	0,06	***	0,09	0,06		0,07	0,07		0,07	0,06	
Citation standing				0,058	0,014	***	0,12	0,01	***	0,12	0,01	***	0,12	0,01	***
Local competitors				0,060	0,037		0,061	0,037		0,057	0,037		0,034	0,036	
Gender: Male vs Female				-0,04	0,07		0,02	0,06		0,02	0,06		0,03	0,06	
Professional age				0,52	0,05	***	0,28	0,05	***	0,27	0,05	***	0,19	0,05	***
Habilitation 2011							0,14	0,06	*	0,13	0,07				
RDD 2011							0,06	0,01	***	0,06	0,01	***	0,06	0,01	***
RDD 2011 2							0,012	0,003	***	0,012	0,003	***	0,011	0,003	***
RDD 2011 3							0,0010	0,0003	**	0,0009	0,0003	**	0,0008	0,0003	**
Hab 2011*researcher vs full p.										0,170	0,067	*			
Hab 2011*associate vs full p.										0,06	0,07				
Hab 2011*Citation standing										-0,044	0,016	**			
Hab 2011*Genetics													-0,09	0,07	
Hab 2011*Managerial Engineering													0,61	0,08	***
Hab 2011*Psychiatry													0,10	0,07	
Hab 2011*Applied Economics													0,26	0,11	*
Group level variance	0,12	0,03		0,11	0,03		0,11	0,03		0,11	0,03		0,12	0,03	
Individual level variance	0,44	0,03		0,52	0,04		0,43	0,03		0,43	0,03		0,41	0,03	
Year level variance	0,46	0,02		0,42	0,02		0,40	0,01		0,40	0,01		0,38	0,01	
Units: group	169			169			169			169			169		
Units: scholar	886			886			886			886			886		
Units: year	5780			5780			5780			5780			5780		

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

The capability of the regression models to predict observed values, can be evaluated by predicted and actual values on several intervals of self-citations, namely: no self-citations, at least one self-citation, above median, mean and 90th percentile of number of self-citations per year (Table 5). The model sensitivity is rather good, as it is able to identify 69% of observations with no self-citations, 89% of a observations with at least on citation, as well as identifying observations above the median (90%) , the mean (72%) and even above 90th percentile (63%). Good predicting performance emerge also considering specificity, positive predictive value and negative predictive value indexes of statistical performance.

Table 5 - Model 4: measure of statistical performance

	no self-citations	self-citations (>0)	> median (2)	> mean (6,33)	> 90th percentile
Sensitivity	69%	89%	90%	72%	63%
Specificity	89%	69%	79%	96%	99%
Positive predictive value	72%	87%	77%	86%	83%
Negative predictive value	87%	72%	91%	91%	96%

sensitivity or true positive rate $TPR = TP / P = TP / (TP + FN)$

specificity (SPC) or true negative rate $SPC = TN / N = TN / (TN + FP)$

precision or positive predictive value $PPV = TP / (TP + FP)$

negative predictive value $NPV = TN / (TN + FN)$

where TP: True Positive; FP: False Positive; TN: True Negative; FN: False Negative

The ratio between the actual number of self-citations and the predicted number of self-citations provides an indication of how frequently scientists over-cite their own work. Predictions are computed according to a Model including baseline, control, and regression discontinuity design variables, while discounting for the effect of the habilitation procedure. Table 6 compares the frequency of observations (author per year) of over and under self-citations along the four sectors, ranks and the periods pre and post habilitation. In the pre-habilitation period (2002-2010), over self-citations account for 43% of the observations while in the post-habilitation period they account for 53% of the observations (+23%). The most remarkable increase is observed among researchers in Economic and Managerial Engineering, from 26% to 59% of the cases (+122%), and in Applied Economics, from 22% to 46% (+102%), as well as among associate professors in Economic and Managerial Engineering, from 35% to 63% (+78%). Therefore, it appears that among these categories of scientists the increase of self-citations has involved a large share of them, and possibly the same self-citation behaviour has changed.

Table 6 – Actual versus predicted self-citations: over and under self-citations observations

		pre-habilitation 2002-2010		post-habilitation 2011-2014		pre-habilitation 2002-2010		post-habilitation 2011-2014		% increase in over self- citation observations
over self-citations:		no	yes	no	yes	no	yes	no	yes	
Genetics	Full Professors	165	195	60	73	46%	54%	45%	55%	1,3%
	Researchers	252	249	134	154	50%	50%	47%	53%	7,6%
	Associate Professors	216	226	90	87	49%	51%	51%	49%	-3,9%
Economic and Managerial Engineering	Full Professors	162	97	73	86	63%	37%	46%	54%	44,4%
	Researchers	203	73	106	151	74%	26%	41%	59%	122,1%
	Associate Professors	175	96	68	116	65%	35%	37%	63%	78,0%
Psychology	Full Professors	222	179	81	78	55%	45%	51%	49%	9,9%
	Researchers	324	204	156	145	61%	39%	52%	48%	24,7%
	Associate Professors	184	136	67	79	58%	43%	46%	54%	27,3%
Applied Economics	Full Professors	63	45	41	39	58%	42%	51%	49%	17,0%
	Researchers	63	18	46	39	78%	22%	54%	46%	106,5%
	Associate Professors	89	66	55	54	57%	43%	50%	50%	16,3%
total observations		2.118	1.584	977	1.101	57%	43%	47%	53%	23,8%

Notes: over self-citations if actual number of self-citations of an author x in a year is above the number of self-citations predicted. To compute the predicted values, we employed MLwiN software, which gives the possibility to first run the model 3 and second to compute predictions excluding the effect of the habilitation variable

4.3 – Robustness test

Since the habilitation effect is rather different across the considered disciplines, four distinct regressions are run as robustness tests. The tests highlight some differences across the considered disciplines in the factors affecting propensity to self-cite and the increase of self-citations following the habilitation.

The importance of group level variance change considerable: it is negligible in Generics, but considerable in the other sectors. Genetics also differs for researchers and associate professors tend to self-cite less than full professors. The professional age is only significant (and strong) in applied economics.

As to self-citation change, the results show that the strategic increase of self-citations is much stronger among researchers when compared to full-professors in Economic and Managerial Engineering (+59%), in Psychology (+39%) and especially in Applied Economics (+106%), while the different is not significant in Genetics. Scientists that were better performing in terms of citations tended to increase self-citations less than their peers, particularly in Genetics and Psychology.

Table 7 - Regression models by recruitment sector

Dependent variables	Genetics			Economic and Managerial Engineering			Psychology			Applied Economics		
	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign
Cons	1,09	0,32	***	-1,79	0,36	***	-0,65	0,32	*	-2,60	0,54	***
Publications year	0,25	0,01	***	0,35	0,01	***	0,12	0,01	***	0,25	0,02	***
Cumulative production year t-1	-0,002	0,003		0,05	0,01	***	0,012	0,002	***	0,05	0,02	**
Square: Cumul. prod. year t-1	0,000018	0,000008	*	-0,000046	0,000014	***	-0,000027	0,000005	***	-0,00041	0,00016	**
Mean coauthors	-0,025	0,006	***	-0,024	0,022		0,011	0,010		-0,055	0,042	
Researcher vs Full Professor	-0,54	0,13	***	0,12	0,19		0,00	0,18		-0,18	0,36	
Associate vs Full Professor	-0,23	0,10	*	0,16	0,14		0,10	0,13		0,32	0,22	
Citation standing	0,18	0,04	***	0,10	0,03	***	0,10	0,03	***	0,04	0,04	
Local competitors	0,03	0,05		-0,02	0,07		0,03	0,07		-0,14	0,16	
Gender: Male vs Female	-0,02	0,08		0,01	0,11		0,07	0,16		0,08	0,21	
Professional age	-0,01	0,09		0,05	0,10		-0,06	0,10		0,60	0,16	***
Habilitation 2011	0,09	0,11		0,32	0,17		0,13	0,13		-0,22	0,28	
RDD 2011	0,042	0,018	*	0,041	0,032		0,045	0,025		0,048	0,052	
RDD 2011_2	0,004	0,004		0,020	0,006	**	0,013	0,005	*	0,020	0,011	*
RDD 2011_3	0,0002	0,0004		0,0018	0,0009		0,0013	0,0006	*	0,0021	0,0015	
Hab 2011*researcher vs full p.	0,109	0,094		0,46	0,16	**	0,33	0,13	*	0,72	0,33	*
Hab 2011*associate vs full p.	-0,03	0,09		0,11	0,15		0,25	0,14		-0,01	0,24	
Hab 2011*Citation standing	-0,086	0,039	*	-0,069	0,039		-0,107	0,027	***	0,028	0,040	
Group level variance	0,05	0,03		0,25	0,09		0,17	0,08		0,30	0,15	
Individual level variance	0,26	0,03		0,18	0,04		0,69	0,09		0,33	0,12	
Year level variance	0,24	0,02		0,36	0,04		0,45	0,03		0,36	0,07	
Units: group	42			36			38			53		
Units: scholar	260			221			258			147		
Units: year	1901			1406			1855			618		

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

Discussion and conclusions

The empirical analysis supports the hypotheses that scientists respond to the introduction of metrics to guide decisions on academic careers by increasing self-citations, particularly those under stronger pressure, namely researchers and scientists with less citations compared to their peers, and scientists in the social sciences.

Some limitations should be discussed. First, some observers may regard inflating self-citations as a minor infringement. However, self-citations have a very important impact on visibility (Fowler and Aksnes, 2007), inflating self-citations is a post-production malpractice that may undermine trust in the publication system (Biagioli, 2016) as well as confidence in the self-policing capability of the academic community. Second, we assumed that a rapid increase in self-citations corresponding to the introduction of the habilitation procedure is likely indicative of a strategic use of self-citations, rather than publications becoming more similar in their focus. Future studies may try to disentangle the two effects. In the mid and long run, it is indeed possible that metric-based evaluation will affect the balance between exploration and exploitation in scientists' research strategies, e.g. towards more specialization and/or small incremental change from one publication to another, thus indirectly leading to more self-citations. As a matter of fact, the more academic rewards become connected to metrics -rather than discovery achievement- the more research and publication strategies are likely to become intertwined. Third, we considered only one national system. This choice was commanded by the fact that Italy represents a valuable case as a forerunner in adopting of a system-wide, metric-based procedure for decisions on career advancement. With these limitations in mind, the results appear to be rather robust.

While analyses of the impact of metrics have tended to focus on the system level, by studying individual change and adopting a multilevel perspective, we avoided the risk of ecological fallacy and we were able to highlight the influence of factors at multiple levels. These findings have implications for the theoretical understanding of scientists' behaviour. First, scientists emerge to be extremely responsive to incentives. In fact, self-citations have increased early from the law on habilitation was issued and well before the first habilitation round. Scientists also appear to be keen to adopt a questionable practice like increasing the number of citations. Relevant variations emerge between disciplines both in the factors predicting the propensity to self-cite, as well as change in self-citations. Significant variability in self-citation is observed not only in relation to the discipline of belonging, but also to the group of belonging, backing the importance of meso-level norms, habits, and competitive pressure. Finally, individual traits, like the professional seniority, as well as the specific pressures to which each individual researcher is subjected, e.g. in terms of competition for career, influence the propensity to self-cite and change in self-citation.

Policy implications emerge in relation to the increasingly wide, pervasive and formalized use of metrics. Indeed indicators of productivity may be valuable in assisting evaluation, and they can be helpful to increase accountability in contexts characterized by lack of meritocracy. However, this study suggests that if not properly designed, metrics can favour the *widespread* diffusion of questionable practices. Therefore, metrics designers and evaluators should carefully take into account the strategic nature of scientists. Among others, this implies discounting self-citations, normalizing citations to avoid citations rings - for instance counting differently if 50 citations are received from two or twenty scientists.

In this article, we focused our attention to the use of self-citations, as they provide a rather fast feedback. In the long run, however, we warn that other unintended effects may occur, that should be taken under observation. In particular, scientists may be gradually abandoning subfields that

are less cited and more challenging in favour of fields that highly cited and less selective. Future research can explore whether the increase in self-citation is durable, or it will retreat once that the pressure and need to self-cite disappear.

References

- Aagaard, K., Bloch, C., Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation*, rvv003.
- Aksnes, D. (2003). A macro study of self-citation. *Scientometrics*, 56(2), 235-246.
- Aksnes, D. W. (2006). Citation rates and perceptions of scientific contribution. *Journal of the American Society for Information Science and Technology*, 57(2), 169-185.
- Aksnes, D. W., and Rip, A. (2009). Researchers' perceptions of citations. *Research Policy*, 38(6), 895-905.
- Anderson, M. S., Louis, K. S., and Jason, E. (1994). "Disciplinary and Departmental Effects on Observations of Faculty and Graduate Student Misconduct." *The Journal of Higher Education*, 65(3, Special Issue: Perspectives on Research Misconduct), 331-350.
- Anderson, M. S., Ronning, E. A., De Vries, R., and Martinson, B. C. (2007). "The Perverse Effects of Competition on Scientists' Work and Relationships." *Science and Engineering Ethics*, 13(4), 437-461.
- Azoulay, P., Graff Zivin, J. S., and Manso, G. (2011). "Incentives and creativity: evidence from the academic life sciences." *The Rand journal of economics*, 42(3), 527-554.
- Bedeian, A. G., Taylor, S. G., and Miller, A. N. (2010). "Management Science on the Credibility Bubble: Cardinal Sins and Various Misdemeanors." *Academy of Management Learning Education*, 9(4), 715-725.
- Biagioli, M. (2016). Watch out for cheats in citation game. *Nature*, 535(7611), 201-201.
- Bohannon, J. (2013). "Who's Afraid of Peer Review?" *Science*, 342(6154), 60-65.
- Bornmann, L. (2013). "Research Misconduct—Definitions, Manifestations and Extent." *Publications*, 1(3), 87-98.
- Bornmann, L., and Daniel, H.-D. (2007). "Multiple publication on a single research study: Does it pay? The influence of number of research articles on total citation counts in biomedicine." *Journal of the American Society for Information Science and Technology*, 58(8), 1100-1107.
- Bush, V. (1945). Science: The endless frontier. *Transactions of the Kansas Academy of Science* (1903-), 48(3), 231-264.
- Clark, B. (1983) *The Higher Education System. Academic Organization in Cross-National Perspective*. Berkeley and Los Angeles: University of California Press
- Cole, J. R., and Cole, S. (1974). Social stratification in science. *American Journal of Physics*, 42(10), 923-924.
- Costas, R., and Bordons, M. a. (2007). "The h-index: Advantages, limitations and its relation with other bibliometric indicators at the micro level." *The Hirsch Index*, 1(3), 193-203.

- de Rijcke, S., Wouters, P. F., Rushforth, A. D., Franssen, T. P., Hammarfelt, B. (2015). Evaluation practices and effects of indicator use—a literature review. *Research Evaluation*, rvv038.
- De Rond, M., and Miller, A. N. (2005). "Publish or Perish: Bane or Boon of Academic Life?" *Journal of Management Inquiry*, 14(4), 321-329.
- Durante, R., Labartino, G., Perotti, R. (2011). Academic dynasties: decentralization and familism in the Italian academia (No. w17572). National Bureau of Economic Research.
- Eiduson, B. T. (1962). *Scientists: Their psychological world*.
- Edwards, M. A., Roy, S. Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science*.
- Engqvist, L., and Frommen, J. (2008). The h-index and self-citations. *Proc. Natl. Acad. Sci. USA*, 99, 11270-11274.
- Fanelli, D. (2012). "Negative results are disappearing from most disciplines and countries." *Scientometrics*, 90(3), 891-904.
- Fowler, J. H., Aksnes, D. W. (2007). Does self-citation pay?. *Scientometrics*, 72(3), 427-437.
- Franzoni, C., Scellato, G., and Stephan, P. (2011). "Changing Incentives to Publish." *Science*, 333, 702-703.
- Garfield, E. (2003). "The meaning of the Impact Factor." *International Journal of Clinical and Health Psychology*, 3(2), 363-369.
- Gelman, A., Imbens, G. (2014). Why high-order polynomials should not be used in regression discontinuity designs (No. 20405). Cambridge, MA: National Bureau of Economic Research. three variables centred in the year of event
- Gianoli, E., Molina-Montenegro, M. A. (2009). Insights into the relationship between the h-index and self-citations. *Journal of the American Society for Information Science and Technology*, 60(6), 1283-1285.
- Giglioli, P.P. (1979) *Baroni e Burocrati*. Bologna. Il Mulino
- Glänzel, W., Thijs, B. (2004). The influence of author self-citations on bibliometric macro indicators. *Scientometrics*, 59(3), 281-310.
- Hackett, E. J. (1994). A social control perspective on scientific misconduct. *The Journal of higher education*, 242-260.
- Hagstrom, W. O. (1965). *The scientific community* (Vol. 304). New York: Basic books.
- Hagstrom, W. O. (1974). Competition in science. *American sociological review*, 1-18.
- Hammarfelt, B., de Rijcke, S. (2014). Accountability in context: effects of research evaluation systems on publication practices, disciplinary norms, and individual working routines in the faculty of Arts at Uppsala University. *Research Evaluation*, rvu029
- Hammarfelt, B., Nelhans, G., Eklund, P., Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, rvv040.
- Harzing, A. W. (2010). *The publish or perish book*. Melbourne: Tarma software research.

- Haustein, S., Larivière, V. (2015). The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. In *Incentives and performance* (pp. 121-139). Springer International Publishing.
- Heinze, T., Shapira, P., Rogers, J. D., and Senker, J. M. (2009). "Organizational and institutional influences on creativity in scientific research." *Research Policy*, 38(4), 610-623.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, 16569-16572.
- Honig, B., and Bedi, A. (2012). "The Fox in the Hen House: A Critical Examination of Plagiarism Among Members of the Academy of Management." *Academy of Management Learning Education*, 11(1), 101-123.
- Hutson, S. R. (2006). Self-citation in archaeology: Age, gender, prestige, and the self. *Journal of Archaeological Method and Theory*, 13(1), 1-18.
- Imbens, G. W., and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.
- Jefferson, T. (1998). "Redundant publication in biomedical sciences: Scientific misconduct or necessity?" *Science and Engineering Ethics*, 4, 135-140.
- Kaboolian, L. (1998). The new public management: Challenging the boundaries of the management vs. administration debate. *Public Administration Review*, 189-193.
- Keizer, K., Lindenberg, S., Steg, L. (2008). The spreading of disorder. *Science*, 322(5908), 1681-1685.
- Kerr, S. (1975). On the folly of rewarding A, while hoping for B. *Academy of Management journal*, 18(4), 769-783.
- King, M. M., Bergstrom, C. T., Correll, S. J., Jacquet, J., West, J. D. (2016). Men set their own cites high: Gender and self-citation across fields and over time. *arXiv preprint arXiv:1607.00376*.
- Leahey, E. (2007). "Not by Productivity Alone: How Visibility and Specialization Contribute to Academic Earnings." *American Sociological Review*, 72(4), 533-561.
- Levin, S. G., Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *The American Economic Review*, 114-132.
- Lissoni, F., Mairesse, J., Montobbio, F., and Pezzoni, M. (2011). "Scientific productivity and academic promotion: a study on French and Italian physicists." *Industrial and Corporate Change*, 20(1), 253-294.
- Lozano, G. A., Larivière, V., and Gingras, Y. (2012). "The weakening relationship between the impact factor and papers' citations in the digital age." *Journal of the American Society for Information Science and Technology*, 63(11), 2140-2145.
- MacRoberts, M. H., MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for information Science*, 40(5), 342.
- Maliniak, D., Powers, R., Walter, B. F. (2013). The gender citation gap in international relations. *International Organization*, 67(04), 889-922.
- March, J. G. (1991). "Exploration and Exploitation in Organizational Learning." *Organization Science*, 2(1), 71-87.

- Martin, B. R. (2007). "Keeping plagiarism at bay. A salutary tale." *Research Policy*, 36(7), 905-911.
- Martin, B. R. (2012). "Does peer review work as a self-policing mechanism in preventing misconduct: a case study of a serial plagiarist", in T. Mayer and N. Steneck, (eds.), *Promoting Research Integrity in a Global Environment*. London/Singapore: World Scientific Publishing/Imperial College Press, pp. 97-114.
- Martin, B. R. (2013). "Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment." *Research Policy*, 42(5), 1005-1014.
- Martinson, B. C., Anderson, M. S., and de Vries, R. (2005). "Scientists behaving badly." *Nature*, 435, 737-738.
- Merton, R. K. (1973). The normative structure of science. In "The sociology of science: Theoretical and empirical investigations", University of Chicago Press.
- Merton RK (1988) The Matthew effect in science, II. Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79:606-623
- Miller, C. C., Glick, W. H., and Cardinal, L. B. (2005). "The allocation of prestigious positions in organizational science: accumulative advantage, sponsored mobility, and contest mobility." *Journal of Organizational Behavior*, 26(5), 489-516.
- Miller, D. (2007). "Paradigm prison, or in praise of atheoretic research." *Strategic Organization*, 5(2), 177-184.
- Moed, H. F. (2000). "Bibliometric indicators reflect publication and management strategies." *Scientometrics*, 47(2), 323-346.
- Moed, H. F. (2005). Citation analysis of scientific journals and journal impact measures. *Current Science*, 89(12).
- Moed, H. F. (2008). UK Research Assessment Exercises: Informed judgments on research quality or quantity?. *Scientometrics*, 74(1), 153-161.
- Morano Foadi, S. (2006). Key issues and causes of the Italian brain drain. *Innovation*, 19(2), 209-223.
- Reale, E., Seeber, M. (2011). Organisation response to institutional pressures in Higher Education: the important role of the disciplines. *Higher Education*, 61(1), 1-22.
- Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, 38(2), 337-341.
- Saha S., S. S., Christakis D. A. (2003). "Impact factor: a valid measure of journal quality?" *JMLA*(91), 42-46.
- Schreiber, M. (2007). Self-citation corrections for the Hirsch index. *EPL (Europhysics Letters)*, 78(3), 30002.
- Snijders, T. B., Bosker, R. R.(2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Second edition. London etc.: Sage Publishers
- Stephan, P. E. (1996). "The Economics of Science." *Journal of Economic Literature*, 34(3), 1199-1235.
- Stephan, P. (2012). Perverse incentives. *Nature*, 484(7392), 29-31.

- Stroebe, W., Postmes, T., and Spears, R. (2012). "Scientific Misconduct and the Myth of Self-Correction in Science." *Perspectives on Psychological Science*, 7(6), 670-688.
- Vandenabeele, W. (2007). Toward a public administration theory of public service motivation: An institutional approach. *Public management review*, 9(4), 545-556.
- Van Dalen, H. P., and Klammer, A. (2005). "Is Science A Case of Wasteful Competition?" *Kyklos*, 58(3), 395-414.
- Van Fleet, D. D., McWilliams, A., and Siegel, D. S. (2000). "A Theoretical and Empirical Analysis of Journal Rankings: The Case of Formal Lists." *Journal of Management*, 26(5), 839-861.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365-391.
- Weingart, P. (2005). "Impact of bibliometric upon the science system: Inadvertent consequences?" *Scientometrics*, 62(1), 117-131.
- Whitley, R. (2007), 'Changing Governance of the Public Sciences', in R. Whitley and J. Gläser (eds) *The Changing Governance of the Sciences. Sociology of the Sciences Yearbook*, pp. 3-27. Dordrecht: Springer.
- Wildgaard, L., Schneider, J. W., Larsen, B. (2014). A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*, 101(1), 125-158.
- Wilson, J. Q., Kelling, G. L. (1982). Broken windows. *Critical issues in policing: Contemporary readings*, 395-407.
- Zuckerman, H. (1977). *Scientific elite: Nobel laureates in the United States*. Transaction Publishers.
- Zuckerman, H (1987) Citation analysis and the complex problem of intellectual influence. *Scientometrics* 12:329-338

APPENDIX – Robustness test: different habilitation years

Dependent variables	Model: Habilitation 2011			Model: Habilitation 2012			Model: Habilitation 2013		
	coeff.	s.e.	sign	coeff.	s.e.	sign	coeff.	s.e.	sign
Cons	-0,32	0,21		-0,18	0,21		-0,01	0,22	
sector: Managerial Engineering	-0,81	0,14	***	-0,82	0,14	***	-0,82	0,14	***
sector: Psychiatry	-1,04	0,11	***	-1,04	0,11	***	-1,04	0,11	***
sector: Applied Economics	-1,25	0,15	***	-1,34	0,15	***	-1,26	0,15	***
Publications year	0,18	0,00	***	0,18	0,00	***	0,18	0,00	***
Cumulative production year t-1	0,01	0,00	***	0,01	0,00	***	0,01	0,00	***
Square: Cumul. prod. year t-1	0,00	0,00	***	0,00	0,00	***	0,00	0,00	***
Mean coauthors	-0,01	0,01	**	-0,01	0,01	**	-0,01	0,01	**
Researcher vs Full Professor	-0,01	0,08		-0,01	0,08		-0,01	0,08	
Associate vs Full Professor	0,09	0,06		0,09	0,06		0,09	0,06	
Citation standing	0,12	0,01	***	0,12	0,01	***	0,12	0,01	***
Local competitors	0,06	0,04		0,06	0,04		0,06	0,04	
Gender: Male vs Female	0,02	0,06		0,02	0,06		0,02	0,06	
Professional age	0,28	0,05	***	0,27	0,05	***	0,27	0,05	***
Habilitation 2011	0,14	0,06	*						
RDD 2011	0,06	0,01	***						
RDD 2011 2	0,012	0,003	***						
RDD 2011 3	0,0010	0,0003	**						
Habilitation 2012				0,10	0,06				
RDD 2012				0,09	0,02	***			
RDD 2012 2				0,012	0,004	**			
RDD 2012 3				0,0006	0,0003	*			
Habilitation 2013							0,00	0,07	
RDD 2013							0,14	0,03	***
RDD 2013 2							0,017	0,007	*
RDD 2013 3							0,0007	0,0004	
Group level variance	0,11	0,03		0,11	0,03		0,11	0,03	
Individual level variance	0,43	0,03		0,43	0,03		0,43	0,03	
Units: year	0,40	0,01		0,40	0,01		0,40	0,01	
Units: group	169			169			169		
Units: scholar	886			886			886		
Units: year	5780			5780			5780		

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