Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca

anvur

National Agency for the Evaluation of Universities and Research Institutes

Research assessment in Italy: have the results of universities converged?

Daniele Checchi, Irene Mazzotta and Sandro Momigliano

National Agency for the Evaluation of Universities and Research Institutions (ANVUR) ROMA, Italy

www.anvur.it

Science, Technology and Innovation indicators (STI 2017). September 6-8, 2017 - Paris, France

Aim of the study and results in a nutshell

- **Performance based funding** are often subject to the criticism that they may produce **cumulative vicious cycles**, since worse performing institutions receive less money, making it more and more difficult to catch up better performing ones.
- In this work we provide **first evidence** on this issue, comparing the results achieved by Italian universities in the two national evaluation exercises, which took place in Italy in 2012 and 2016.
- After making the two exercises comparable, we find that the **dispersion in research quality of the Italian universities falls in the second exercise**. Convergence is largely due to changes in relative performance of researchers participating to both exercises and to the hiring decisions of universities.

Context: a simplified overview of the two Italian research assessment exercises (1)

Both exercises:

- followed the same approach based on a combination of peer review and bibliometric methods. They referred, respectively, to the years 2004-2010 and 2011-14.
- Evaluation was carried out by panels of experts (about 450 experts in each exercise, with 61 who participated to both) who recruited about 15,000 external referees in each exercise
- were highly publicized, making a noteworthy impact on the reputation of institutions involved
- had a direct impact on the public funding of the universities, based on the average performance of their researchers' staff. Compared to a proportional distribution, the scheme modified allocation of total public funds to individual institutions by up to 4.5% (on average by 2%) in the second exercise.

Context: a simplified overview of the two Italian research assessment exercises (2)

- Here we focus on the main performance index, which applied to all researchers in the universities and guided the allocation of around 13% of public funds.

The share of funds z going to institution j is based on the following formula:

$$z_j = \frac{\sum_{i=1}^{n_j} v_{ji}}{\sum_{j=1}^{k}\sum_{i=1}^{n_j} v_{ji}} = \frac{\bar{v}_j}{\bar{v}} \cdot \frac{n_j}{n}; \quad \sum_{j=1}^{k} z_j = 1, \quad \sum_{j=1}^{k} n_j = n$$

where **v_{ji} is the mark obtained by researcher *i* in institution *j***; k is the number of institutions and n_j the size of research staff in institution j.

That is, if the institution average $\bar{v}_j$ was equal to the national average $\bar{v}$, the fund share corresponded to its share of the research staff at the national level n_j/n; given a staff share, the higher is the research performance, the larger will be the funds obtained.

Basic data

The universities that participated to the exercises had very different sizes, in terms of researchers staff.

Quartiles	VQR 2004-2010		VQR 2011-2014	
	# universities	# researchers	# universities	# researchers
1st quartile	22	35,415	22	33,188
2nd quartile	23	14,075	23	13,540
3rd quartile	23	4,645	23	4,753
4th quartile	23	823	23	1,095
Total	**91**	**54,958**	**91**	**52,576**

Harmonization of grades

Research product allocation	VQR 2004-2010	VQR 2011-2014	Harmonisation
Fraud	(self) plagiarism: -2	not assessable: 0	limited: 0
Wrong deliverables	not assessable: -1	not assessable: 0	
Absence of deliverables	missing: -0.5	not assessable: 0	
Decile* 1	limited: 0	limited: 0	
Decile* 2			
Decile* 3		acceptable: 0.1	
Decile* 4			
Decile* 5			
Decile* 6	acceptable: 0.5	fair: 0.4	fair: 0.4
Decile* 7	good: 0.8		
Decile* 8		high: 0.7	high: 0.7
Decile* 9	excellent: 1		
Decile* 10		excellent: 1	excellent: 1
Ex-ante mean score	**0.41**	**0.35**	**0.32**
Ex-ante std. dev. of scores	**0.43**	**0.33**	**0.36**

** Decile in the world distribution of quality*

Distribution of universities' scores in the two research assessment exercises (1)

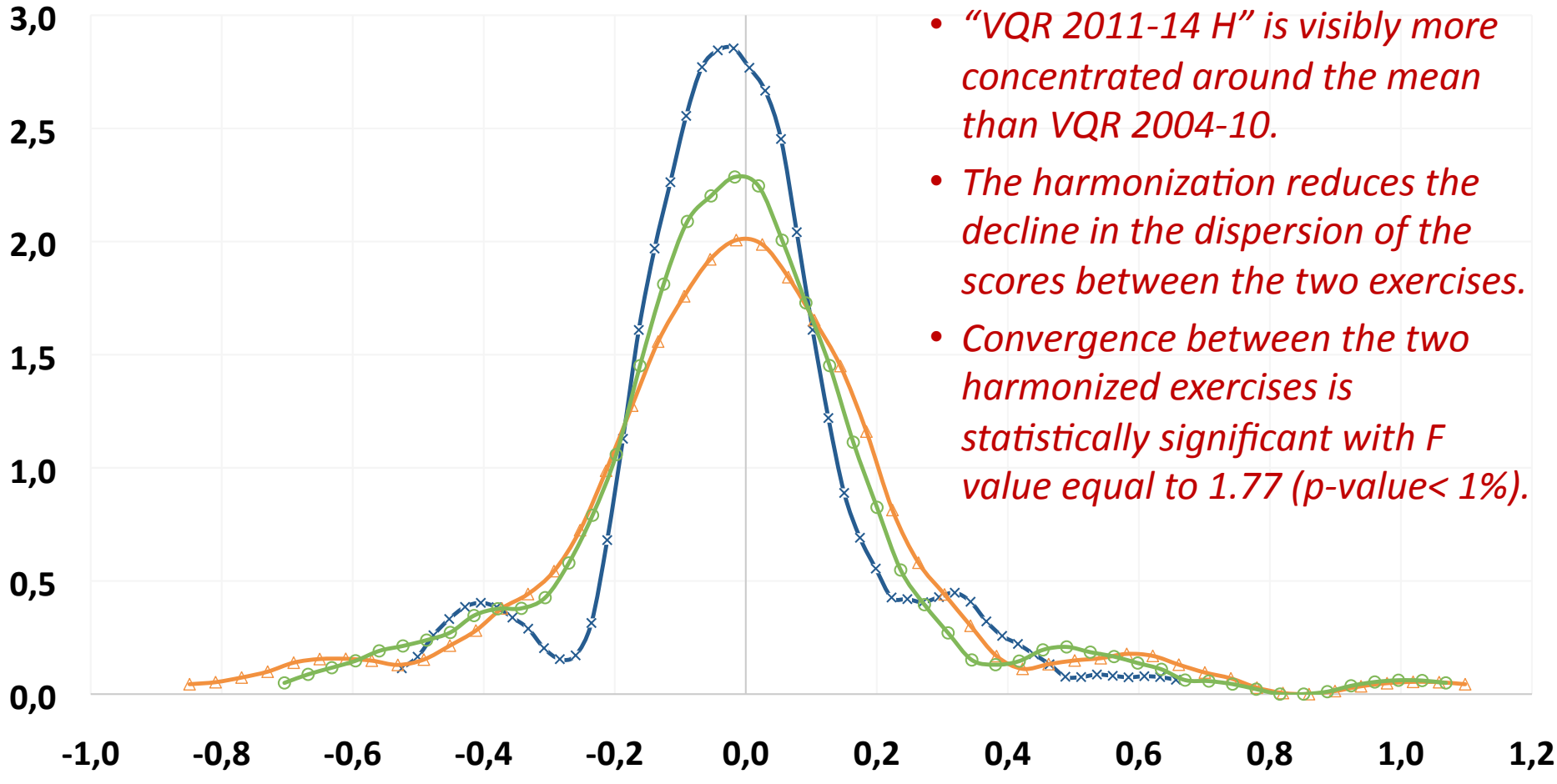
For testing the convergence/divergence of universities in terms of quality of research, we compute the following statistics

$$s_{jt} = \left(\frac{\bar{v}_{jt}}{\bar{v}_t} - 1\right); \qquad j = 1, \ldots, 92; \qquad t = 1,2$$

where s_{jt} is the deviation of university j from the mean grading in the research assessment exercise t.

A positive score indicates that the university obtains an above-mean grading, while a negative value implies a below-mean performance.

Distribution of universities' scores in the two research assessment exercises (2)



- *"VQR 2011-14 H" is visibly more concentrated around the mean than VQR 2004-10.*
- *The harmonization reduces the decline in the dispersion of the scores between the two exercises.*
- *Convergence between the two harmonized exercises is statistically significant with F value equal to 1.77 (p-value< 1%).*

Std. dev.	***0,19***	***0,27***	***0,25***
Weighted Std. dev.	***0,11***	***0,16***	***0,14***

A model for examining the dynamics of universities' grades

$$\Delta s_j = (s_{jt} - s_{jt-1}) = \alpha + \gamma \cdot s_{jt-1} + \epsilon_j$$

where γ measures the dependence from initial conditions:

- $\gamma < 0$ **implies convergence** (i.e. regression to the mean); the closer the estimated γ to -1 the quicker is the convergence
- $\gamma > 0$ (or it is below -2) **implies divergence**

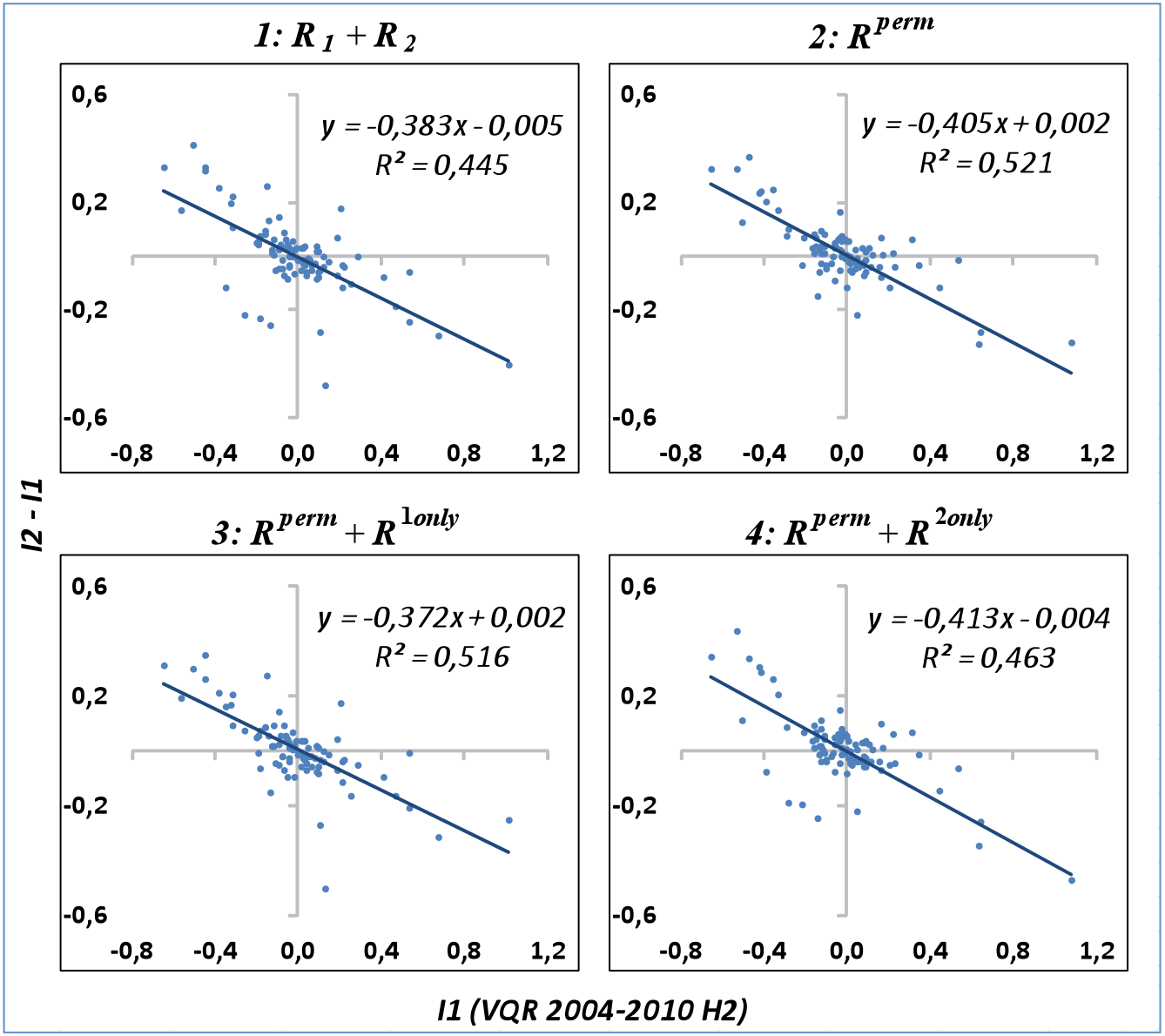
For better understanding the causes of convergence we split the researchers involved in the two exercises (R_1 and R_2, respectively) into subgroups, according to the following decomposition:

$$R^{perm} + R^{1only} = R_1$$

$$R^{perm} + R^{2only} = R_2$$

"perm" → who participated in both exercises with the same academic position
"1only" → who participated only to the first exercise
"2only" → who participated only to the second or were promoted after the first exercise.

Applying the model to 4 different populations (1)



- Highly significant negative regression coefficients

  => CONVERGENCE

- Coefficient of the 4th regression is the largest (in absolute value) => **relatively large contribution to convergence of recruited and promoted** (R^{2only}).

Applying the model to 4 different populations(2)

R_1+R_2: the regression coefficient (-0.383) **shows convergence**. On average, in the second exercise **universities reduce by more than a third their initial distance from the mean grading**.

R^{perm}: the regression coefficient (-0,405) shows a faster convergence to the mean if we consider only those **researchers participating to both exercises** with the same academic position.

R^{perm}+R^{1only}: **including the effects coming from retirement** (largely determined by age), the regression coefficient (-0,372) gets closer to zero.

R^{perm}+R^{2only}: the regression coefficient (-0.413) is the highest (in absolute value), suggesting a **relatively large contribution to the convergence coming from universities decisions on recruitment and promotions**. As the performance related scheme is largely targeted to universities, it is here that we should see the largest impact of the reputation and monetary incentives.

Conclusions and future work (1)

In this paper we look at the performance of individual Italian universities in the two national evaluation exercises of research quality, which took place in Italy in 2012 and 2016.

We find that, contrary to what was expected by critics of performance based funding, in the Italian context there has been a **significant convergence in the quality of research of involved institutions**.

Conclusions and future work (2)

Our results suggest that this convergence was induced by the **monetary and reputation incentives**. Apparently, the fact that worse performing institutions received less money did not hamper their efforts to catch up better performing ones.

These are **preliminary results** and should be taken with caution. They could be the effect of the relatively **small amount of funds shifted by the scheme**; at most 4.5% of public funds received by the individual universities. Also, the convergence may be an **initial** (not permanent) effect of the scheme.

We plan to further study this issue, examining the determinants of the convergence and refining our harmonization of the two exercises.

Thanks for your attention!