

A Doubly Conditional performance evaluation model, the Democratization of evaluation and Altmetrics



SAPIENZA
UNIVERSITÀ DI ROMA

Cinzia Daraio (E-mail:daraio@dis.uniroma1.it)

DIAG Dipartimento di Ingegneria Informatica, Automatica e Gestionale Antonio
Ruberti

STI 2017

Paris, 6-8 September 2017

Outline

- Introduction
- A framework
- A doubly conditional performance evaluation model
- The problem of Democratization of the Evaluation and Altmetrics
- Conclusions

Introduction: an overall framework for the assessment of Research and its impacts

- Crucial importance of the issue of **designing relevant models of indicators** to assess research and its impacts.
- The evaluation of research activities is a **complex task** for many reasons.
- There are **no perfect indicators** or metrics which fit for all purposes.
- In order to **understand the appropriateness** of the indicators to be used, we need to frame the problem taking into account
 - **the systemic nature** of the phenomena and
 - **to develop models of metrics** that are as close as possible to the reality being assessed.

Introduction: an overall framework for the assessment of Research and its impacts

- According to Daraio (2017a), each metric of research assessment is based on a *model* that can be **implicitly** or **explicitly** defined and discussed.
- If the model underlying the metric is not described, this does not mean that the indicator is more robust to modelling choice. It simply means that **you do not explicitly clarify** and account for the underlying theoretical choices, methodological assumptions and data limits.
- Thus, as a consequence, if you do not specify your model of the metric, you **may not check its robustness**.

Introduction: an overall framework for the assessment of Research and its impacts

Developing models is important for two main reasons:

1. **to learn** about the explicit consequences of assumptions, test the assumptions, highlight relevant relations;
2. **to improve**, to better operate, document/verify the assumptions, decompose analysis and synthesis, systematize the problem and the evaluation/choice done, explicit the dependence of the choice to the scenario.

There are however several **pitfalls and difficulties** in modelling, which mainly relate to:

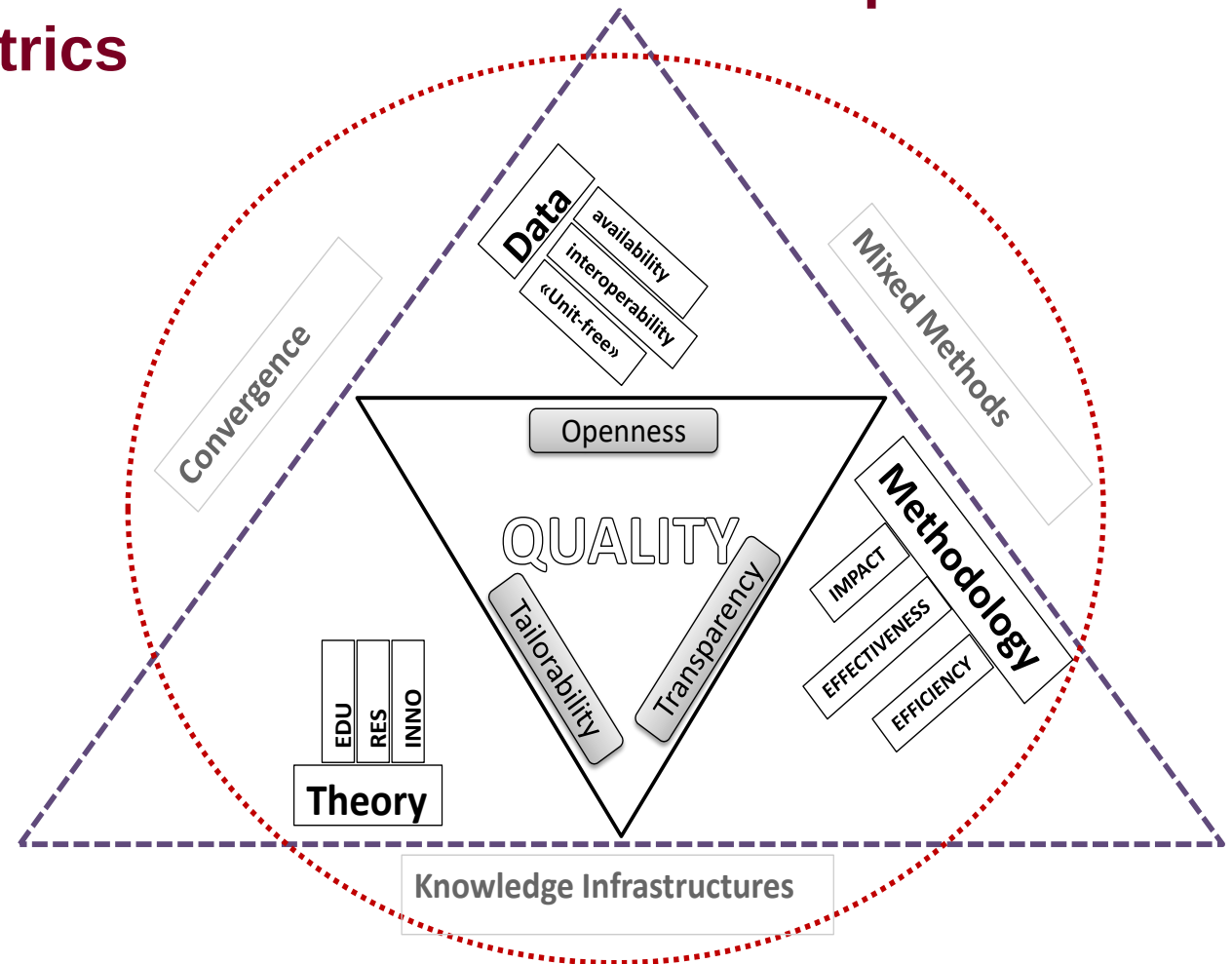
- the possibility that the targets are **not quantifiable**;
- the **complexity, uncertainty and changeability** of the environment in which the controlled system works and,
- the **limits** in the decision context;
- the **intrinsic complexity of calculation**.

Introduction: an overall framework for the assessment of Research and its impacts

Within this context, Daraio (2017a) proposes *a framework*, intended as a background which includes the main conditions, circumstances, ideas, and so on, to the realization of activities related to the assessment of research and its impacts.

It is suggested *as a reference* to develop models of *metrics* accounting for the systemic nature of the research activity and its interrelations with teaching and innovation activities. See Figure 1 .

Fig. 1 A systemic framework for the development of models of metrics

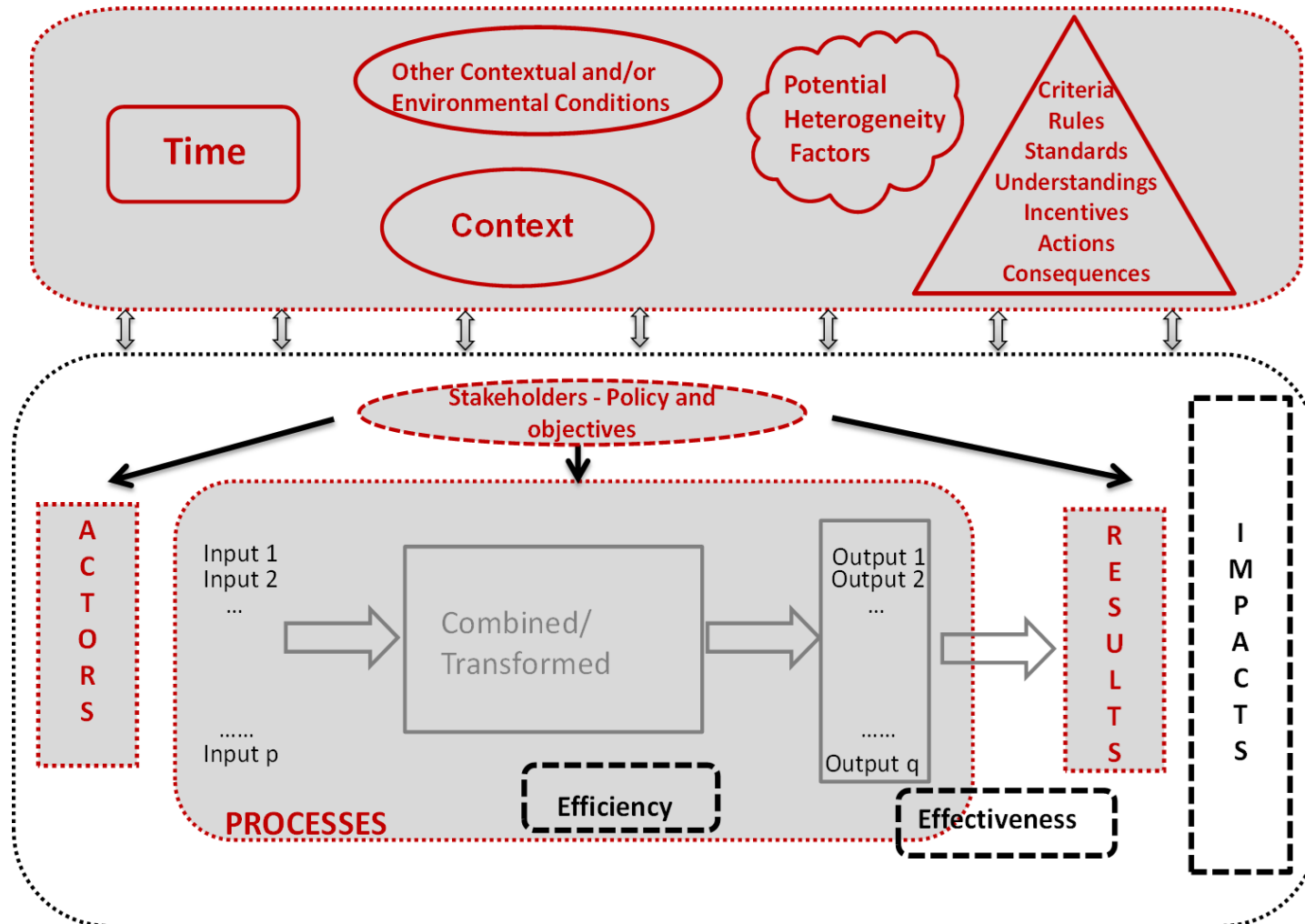


The ability to develop (and afterwards understand and use effectively) models for the assessment of research is linked and depends, among other factors, on the degree or depth of the conceptualization and formalization, in an unambiguous way, of the underlying idea of quality.

A Doubly-Conditional Performance evaluation model

- Figure 2 illustrates the main component of the *Doubly Conditional performance evaluation model* proposed by Daraio (2017b) which is based on a combination and extension of Johnsen (2005); van den Hove (2007) and Lewis (2015).
- It is “doubly conditional” because the evaluation is **conditioned two times: on the information that are available and on those which are not available.**

A Doubly-Conditional Performance evaluation model



A Doubly-Conditional Performance evaluation model

We distinguish two kind of conditioning:

1. **Internal conditioning or *normalization*:** on the items reported in the bottom of Figure 2 (actors, processes and results) means to compare comparable entities, setting appropriate reference sets.
2. **External conditioning or *contextualization*:** on the items reported in the top of Figure 2 equals to account for heterogeneity factors that we call *external conditioning* or contextualization.

According to this model of performance evaluation,

it's all a matter of appropriate normalization and contextualization.

A Doubly-Conditional Performance evaluation model

This model:

1. Permits the *identification of the components* of the analysis (in terms of theory-method-data characterization) that are excluded (what remains outside) in the specific context of the evaluation;
2. Provides an *interpretative value of the measure* (or metrics or indicators) of research assessment calculated, that has to be considered as a *residual*, what remains after the consideration of the dimensions we pursued, that is due to other factors/components not accounted for;
3. Represents *a step toward the democratization of the evaluation practice*, able to balance the opposite views of external accountability and internal improvement (Ewell, 2009).

THE PROBLEM OF “DEMOCRATIZATION” OF THE EVALUATION AND ALTMETRICS

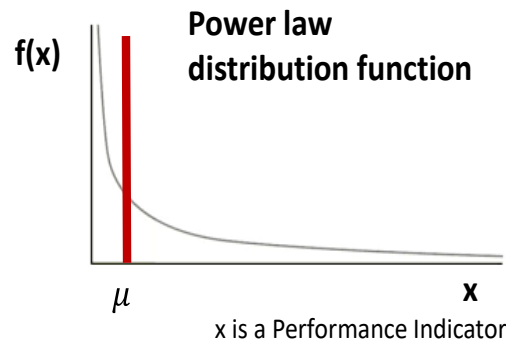
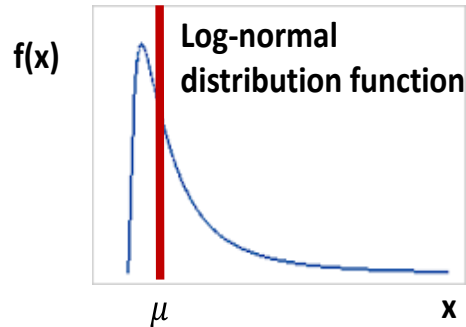
The main contribution of this paper is:

- to propose the *doubly conditional performance evaluation model* as a *democratic evaluation tool* for “value creation” in a learning and participatory environment.
- It may be seen as a *revisited version of the Ricardo’s approach of comparative advantages* but in the context of a broader framework (Figure 1).

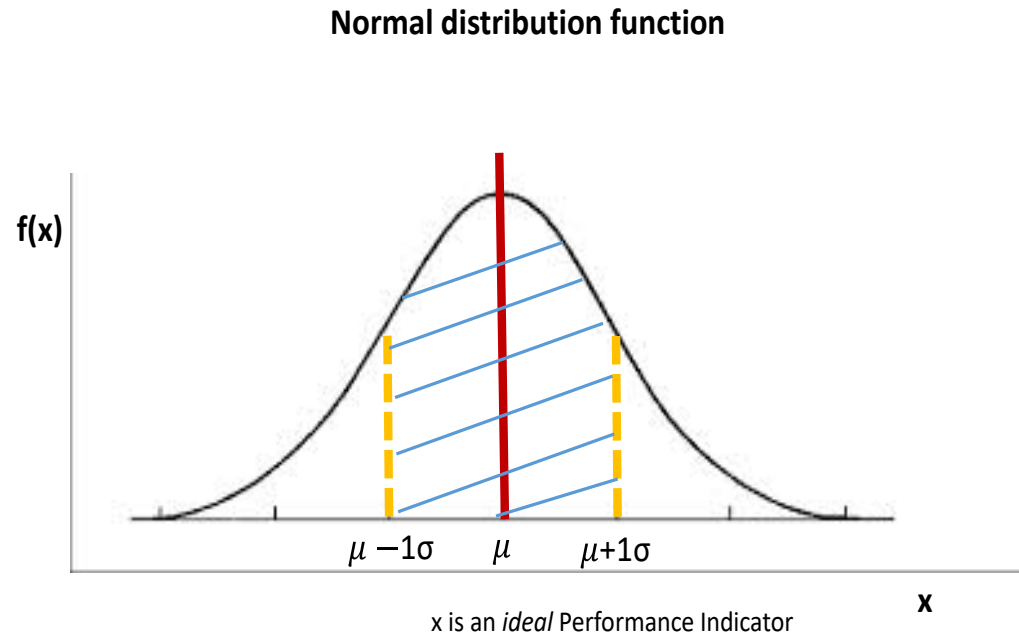
Main conceptual building blocks

Building block	Selected References
1. Complexity of the assessment of research	Need to adopt a systematic view, complexity of the assessment linked to the “the implementation problem”, introduction of the Doubly conditional performance evaluation model (Daraio, 2017b); multidimensionality of the assessment of the research (Moed and Halevi, 2015); problems of data quantification and standardization for different evaluation and assessment purposes (Glänzel, 1996, Daraio and Glänzel, 2016)
2. Recent critiques of traditional bibliometric indicators	Cronin and Sugimoto, (2014, 2015); Dora Declaration, Hicks et al. (2015); Wilsdon (2015)
3. Development of Altmetrics	Priem et al. (2010, 2012), Thelwall et al. (2013), Glänzel, and Gorraiz (2015), Haustein, S., Bowman, T. D., & Costas, R. (2015), Costas, C. R., Haustein, S., Zahedi, Z., & Larivière, V. (2016), Ràfols, I., Robinson-García, N., & van Leeuwen, T. (2017)
4. Meanings of Altmetrics	1) Rousseau and Ye (2013) 2) Altmetrics as the sign of the computerization of the research process: Moed (2016) with link to Nielsen (2012)
5. Towards the democratization of the evaluation of research	Science-policy interfaces (van den Hove, 2007); economics of democracy (Acemoglu and Robinson, 2006), deliberative policy learning (Kowarsch et al. 2016)
6. Examples of calls for “democratization” of metrics	1) Hill (2016): “making impact assessment mainstream”; 2) Douglass (2016): The New Flagship University: Changing the Paradigm from Global Ranking to National Relevancy.

An illustration of the problem of “democratization” of evaluation as need of an answer to the *injustice (unfairness)* generated by the skewness of performance indicators. Our claim: Altmetrics as a possible answer.



The *expected value* μ is reached by only a small fraction of the population.
There are a *few outlying performers* and a long queue.
These $f(x)$ highlight the *inequality* of the population!



The *expected value* μ is reached by half of the population.
In around 68% of the cases, an observation randomly chosen will fall within $\mu \pm 1\sigma$: the majority of the population is around the average performance.
This $f(x)$ highlights the *cohesion* of the population!

An application of the Economics of democracy

The approach we proposed may be:

- a first step towards a formative democratic approach to evaluation in which **indicators** are used as **learning tools** instead of **target of policy**.
- An interesting exercise could be the application of the economic framework proposed by **Acemoglu and Robinson (2006)** for analysing the *creation and consolidation of democracy* in the context of research evaluation.

An application of the Economics of democracy

Acemoglu and Robinson (2006) component	Application in the context of research evaluation
1. the strength of civil society	the movement against the blinded use of bibliometric indicators
2. the structure of political institutions	science-policy interface (see below)
3. the nature of political and economic crises	the crisis of science (Benessia et al. 2016)
4. the level of economic inequality	inequality which comes out from the skewness of the bibliometric indicators
5. the structure of the economy	structure of the sciences and their linkages
6. the form and extent of globalization	form and extent of the globalization in science

Building blocks of deliberative policy learning.

Source: Kowarsch et al. (2016, p.8 Table 3)

Representation	Incorporating wide variety of viewpoints and stakeholders
	Incorporating wide variety of scientific insights and disciplines
Empowerment	Critically scrutinizing access barriers; supporting development of policy options and scenarios reflecting marginalized viewpoints
	Critically scrutinizing requirements to adequately participate; organizational support throughout the process
Capacity building	Building internal capacity of participants: knowledge integration and synthesis
	Realizing external capacity building: providing knowledge about implications of alternative policy pathways, disclosing key uncertainties and normative assumptions
Spaces for deliberation	Realizing continuous and iterative face-to-face deliberation
	Realizing vertical and horizontal linkage of spaces for deliberation

Conclusions

- The critics to traditional bibliometric indicators (constructed on number of publications and citations) is exacerbated by *some unpleasant and tricky properties* these indicators have, e.g. skewness and asymmetry, which translate in highlighting the *inequality (disparities)* among the assessed units.
- The critics of bibliometric indicators have increased over the years, also because, among other factors, there has been an increasing usage of bibliometric indicators at the individual level.
- When indicators are used as metrics in research assessment in which *individuals* are the unit of the analysis, much more care should be given to the *issues of “democratization”* of the tools used.

Conclusions

- A possible application of the *doubly conditional* performance evaluation model described so far is that if we are able to account for all its components, the units of analysis (e.g. scholars, institutions...), i.e. the entities under assessment, may be seen as performing at the best of their possibilities.
- Put in another way, *it exists for each entity an “optimal path” or trajectory in the multidimensional performance framework along which it is top performing, which permits to reach its best possible result.*

Conclusions

- The doubly conditional approach we propose in this paper could help **to develop participation, interaction, and learning** aspects of the units analyzed in a multidimensional evaluation framework **where each one can find its own best way to contribute to the “creation of value”** of its over-hierarchical institution (for example, for individuals the department of the university, for the university the region in which it is located, and so on).
- **Further investigations** are needed to have a more complete view on the important issue of the “democratization” of evaluation.
- In this paper we give **only some starting reflections** that need further discussion and consolidation.

Some Selected references

- N.B. Note that the version of this paper in the STI usb pen was not updated: for receiving the updated version together with the following papers, please write at: daraio@dis.uniroma1.it
- Daraio C. (2017a). A Framework for the Assessment of Research and its Impacts. Forthcoming in *Journal of Data and Information Science*.
- *Daraio C. (2017b). Assessing Research and its Impacts: The Generalized Implementation Problem and a Doubly-Conditional Performance Evaluation Model, accepted paper, ISSI 2017 Conference, Wuhan, China, October 2017.*