

VQR 2011-2014: Note sui criteri adottati dal GEV13 Sub-GEV “Economia e Storia Economica” *

A seguito della pubblicazione della classificazione delle riviste dell'area 13 e i relativi comunicati del GEV 13 (ad oggi: del 14, 22 e 29 gennaio 2016), proponiamo una rapida analisi delle problematiche relative alla non discriminazione nella valutazione delle riviste e delle aree disciplinari, e segnaliamo alcune incongruenze tra quanto richiesto dal MIUR – con il D.M. 458 del 27-06-2015 (“Linee guida VQR 2011-2014”)¹ – e quanto realizzato dal GEV 13.

Come cerchiamo di mostrare, i criteri e la metodologia adottati dal GEV 13 nel contesto dell'attuale VQR 2011-2014 rischiano di minare il pluralismo dei metodi e degli approcci nella ricerca economica, come già è avvenuto in passato,² e di strumentalizzare la valutazione della ricerca a fini ad essa estranei, di politica culturale e accademica.³

Le principali problematiche relative all'operato del GEV 13 riguardano:

- (i) l'uso automatico della bibliometria, pur essendo l'area 13 non bibliometrica, senza peraltro l'adozione delle metodologie usate dai ‘veri’ settori bibliometrici (ad esempio la standardizzazione degli indici bibliometrici per aree tematiche);
- (ii) l'applicazione di criteri di valutazione rigettati dalla letteratura scientometrica, come ad esempio l'utilizzo delle citazioni ricevute dalla rivista come criterio di qualità del singolo articolo;
- (iii) arbitri e incoerenze nei criteri metodologici adottati e nella loro applicazione (ad esempio in relazione alle riviste italiane).

In questo documento, illustriamo questi problemi tramite esempi relativi per lo più alla sotto-area “Storia economica”, per via del numero più contenuto di riviste che ne sono incluse. E' evidente però che si tratta di problemi comuni a tutte le sotto-aree.

Questi problemi possono condurre a diversi risultati distorsivi, sia tra aree e settori scientifico-disciplinari, sia tra riviste. A loro volta, tali distorsioni sono problematiche sia dal punto di vista del dibattito culturale,⁴ sia dal punto di vista della politica universitaria, nel senso delle possibili conseguenze in termini di finanziamento e premialità legati ai futuri risultati della valutazione. In tale ambito più ristretto, alcune delle disparità di trattamento presentano perfino evidenti rischi di generare ricorsi amministrativi o giudiziari. Per ovviare ad almeno alcuni dei problemi rilevati, ci sembra necessario procedere come minimo ai seguenti passi:

- 1) garantire la valutazione in peer review per qualsiasi prodotto ne sia fatta richiesta;
- 2) procedere alla standardizzazione per ambito disciplinare di tutte le misure citazionali e degli indici bibliometrici, e delle valutazioni in peer review;
- 3) utilizzare un mix di indicatori bibliometrici di articoli e di riviste, e chiarire la fonte e il metodo di reperimento dei dati;
- 4) applicare gli stessi criteri e metodi per tutte le riviste e tutti i prodotti, senza discriminazioni;
- 5) garantire ampia trasparenza anche nella peer review, e prendere le opportune misure per evitare che i referee valutino gli articoli sulla base della sede di pubblicazione.

* A cura di Carlo D'Ippoliti e Giulia Zacchia, Sapienza Università di Roma; per commenti: carlo.dippoliti@uniroma1.it. Il sub-GEV “Economia e Storia economica” comprende gli ssd: SECS/P01, SECS/P02, SECS/P03, SECS/P04, SECS/P06, SECS/P12.

¹ http://www.anvur.org/attachments/article/825/dm_27_06_2015_vqr_11-14.pdf

² Ci riferiamo in particolare al dibattito legato al primo esercizio di Valutazione Triennale della Ricerca (VTR 2001-2003) e al *Rapporto Finale* dell'allora GEV 13, in particolare le note critiche del componente del panel Luigi Pasinetti e le relative risposte del presidente del panel Guido Tabellini (http://vtr2006.cineca.it/rel_area/panel_13.pdf). Per una panoramica sull'importanza e l'origine storica del pluralismo di metodi e approcci in economia in Italia, si veda Pasinetti L.L. e Roncaglia A. (2006), “Le scienze umane in Italia: il caso dell'economia politica”, *Rivista Italiana degli Economisti*, vol. 11, pp. 461–499.

³ Per maggiori informazioni si rimanda a Corsi M., D'Ippoliti C. e Lucidi F. (2011), “On the Evaluation of Economic Research: The Case of Italy”, *Economia Politica*, vol. 28, n. 3, pp. 369-402.

⁴ Nel campo dell'economia mainstream, importanti riflessioni critiche sullo stato della scienza economica sono state formulate soprattutto alla luce della recente crisi negli USA; il dibattito ha preso avvio almeno dall'intervento di Krugman (2009), disponibile alla URL http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html?_r=0, e si è fortemente sviluppato dopo di allora.

1. PRINCIPALI LIMITI METODOLOGICI

1.1. L'uso della bibliometria

Il primo problema riguarda l'utilizzo, praticamente esclusivo per quanto riguarda gli articoli su rivista, e di fatto automatico, dei criteri bibliometrici, nonostante l'area 13 non sia formalmente definita come area bibliometrica. Permane dunque la tesi – ormai screditata dalla comunità scientifica internazionale⁵ – che sia possibile “valutare” un lavoro scientifico senza leggerlo. Peraltro, sono considerati solo indici basati direttamente o indirettamente sul numero di citazioni ricevuto da un lavoro, con evidente confusione tra la *qualità* e la *visibilità* di un prodotto di ricerca (dove il DM 458 richiede la valutazione della prima, mentre gli indici considerati misurano solo, al più, la seconda).

Infatti, il GEV13 ha considerato quattro indici citazionali (5-year Impact Factor, di seguito indicato IF5; Article Influence Score, AIS; Impact per Publication, IPP; SCImago Journal Rank, SJR), tratti da due banche dati private (Web of Science (WoS) di Thomson Reuters, per i primi due, e Scopus di Elsevier per i secondi due). Sulla base dell'ordinamento di questi, il GEV ha stilato quattro diversi ranking di riviste (uno per indicatore), a loro volta divisi in 4 sotto-aree: “Economia”, “Storia Economica”, “Economia Aziendale” e “Statistica e metodi matematici per le decisioni”. A tal proposito, occorre notare:

- a) (**esclusività e comparabilità dei metodi**) la decisione di valutare alcuni prodotti (gli articoli su rivista) esclusivamente tramite bibliometria ed altri prodotti (principalmente libri e capitoli di libro) esclusivamente tramite peer review rischia di generare discriminazioni sistematiche e non pienamente quantificabili ex ante. Infatti, empiricamente il grado di concordanza tra valutazioni bibliometriche e valutazioni in peer review è oggetto di discussione in letteratura, e in ogni caso dipende evidentemente dagli specifici criteri bibliometrici adottati. Alla luce dei precedenti casi italiani,⁶ esiste il rischio che la valutazione in peer review conduca a valutazioni sistematicamente più basse della bibliometria.
- b) (**criteri di qualità**) come riportato nel suo documento “Criteri per la valutazione dei prodotti di ricerca”⁷ (indicato da qui in poi con *Criteri*), il GEV 13 considera i quattro indici selezionati come degli indicatori di “impatto” (p. 12). Dunque, per stessa ammissione del GEV, l'uso esclusivo di questi indicatori, non integrato da peer review, configura un'applicazione del DM 458 al più molto parziale, in quanto all'art. 5 il DM recita “Il giudizio di qualità si baserà sulla valutazione della pubblicazione *in base alla sua originalità, rigore metodologico e impatto* attestato o potenziale nella comunità scientifica internazionale” (comma 2, corsivo e sottolineatura aggiunti).
- c) (**ranking di prodotti e di riviste**) la considerazione di soli indicatori di impatto e la redazione di ranking di riviste viola la lettera dello stesso art. 5 del DM 458, che prevede la definizione di fasce

⁵ Per la critica dell'utilizzo del Journal Impact Factor per la valutazione della ricerca, si veda la San Francisco Declaration on Research Assessment (<http://www.ascb.org/dora/>), o più in generale i 10 punti del *Leiden Manifesto* (<http://www.leidenmanifesto.org/>), di cui riportiamo ai fini della nostra analisi in particolare il primo punto: “Quantitative evaluation should support qualitative, expert assessment”, ed il sesto punto: “Account for variation by field in publication and citation practices”. Più in generale, si rimanda agli importanti lavori preparatori del REF (Research Excellence Framework) nel Regno Unito, e in particolare: Wilsdon J. *et al.* (2015), *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*, Higher Education Funding Council for England, Londra, e i relativi allegati: <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463.en.html>

⁶ Si vedano ad es. Bertocchi G. *et al.* (2015), “Bibliometric evaluation vs. informed peer review: Evidence from Italy”, *Research Policy*, vol. 44 n. 2, pp. 451-466; e contra: Baccini A. e De Nicolao G. (2015), “Do they agree? Bibliometric evaluation vs informed peer review in the Italian research assessment exercise”, *arXiv*, 1505.00115 [cs.DL], disponibile alla URL <http://arxiv.org/abs/1505.00115>.

⁷ http://www.anvur.it/attachments/article/856/criteri%20GEV%2013_italiano%200~.pdf

di merito (“livelli di qualità”) *delle pubblicazioni*, non delle riviste, basate sui criteri di originalità e rigore, da cui si assume conseguire o evincere un impatto attuale o futuro. Ad es. la prima fascia (pubblicazione “eccellente”) è definita come: “la pubblicazione raggiunge i massimi livelli in termini di originalità e rigore metodologico, e ha conseguito o è presumibile che consegua un forte impatto nella comunità scientifica di riferimento” (comma 2, lettera a); oppure l’ultima fascia (pubblicazione “limitata”): “la pubblicazione non raggiunge livelli di originalità e rigore metodologico tali da essere considerata significativa dalla comunità scientifica di riferimento” (comma 2, lettera f).

d) (**retroattività dei ranking**) la decisione di procedere alla stesura di un ranking di riviste (anziché un ranking dei prodotti) ai fini della valutazione di lavori pubblicati tra il 2011 e il 2014 lede fortemente il diritto degli enti di conoscere in anticipo il modo in cui verranno valutati e programmare corrispondentemente le proprie attività. Infatti, i ricercatori affiliati a tali enti hanno deciso a quali riviste inviare i propri lavori senza conoscere quale “valore” a tale riviste sarebbe poi stato attribuito.

Inoltre, tale procedura viola la parità di trattamento tra enti, nella misura in cui alcuni enti di ricerca godono ex post del vantaggio di avere unità del proprio personale nel comitato (il GEV 13, in questo caso) cui è demandata la definizione dei criteri di valutazione. Questo punto è particolarmente rilevante in quanto, come mostrato oltre, i ranking fondati sulla bibliometria sono molto sensibili anche a piccole modifiche nei dettagli dei criteri (ad es. di costruzione degli indici e/o di loro confronto e aggregazione).

e) (**fascie di merito: l’universo di riferimento**) per quanto riguarda la classificazione in classi di merito delle riviste, il DM 458 definisce le classi in termini relativi, con riferimento a ipotetici decili della distribuzione di qualità “della distribuzione della produzione scientifica internazionale dell’area cui [il prodotto] appartiene” (art. 5). Questi decili, dunque, sono definiti in relazione alla complessiva produzione economica internazionale, e non con ristretto riferimento al campione di prodotti o riviste considerati dal GEV. La distinzione è molto rilevante, in quanto il campione considerato dal GEV 13 non è un campione casuale, e complessivamente rappresenta l’insieme delle riviste più visibili e prestigiose della “produzione scientifica internazionale dell’area”. Dunque, la definizione delle classi di riviste è illegittima e ingiustificatamente troppo ristretta, con la conseguenza di restringere il numero di riviste classificate come “eccellenti” (che per definizione sono il 10% del “totale”).

Ad ogni modo, è previsto un meccanismo di validazione della valutazione bibliometrica: un campione casuale dei prodotti valutati tramite bibliometria (costituito da almeno il 10% degli articoli su rivista) verrà valutato anche con peer review, per verificare il grado di concordanza tra i due metodi di valutazione. Tuttavia, come riportato nei *Criteri*, in caso di divergenza tra valutazione bibliometrica e peer review, “la valutazione finale degli articoli appartenenti al campione resterà bibliometrica” (p. 14). A tal proposito, occorre notare:

f) (**automatismo della valutazione**) questa scelta porta all’evidente paradosso che uno o più esperti avranno valutato un prodotto, leggendolo, pari a un certo livello qualitativo, e ciononostante per finalità istituzionali il prodotto sarà valutato con metodi automatici (bibliometrici), del tutto indipendenti dal contenuto del lavoro (di nuovo, con evidente confusione delle caratteristiche di *qualità* e di *visibilità* dei prodotti).

g) (**causalità e correlazione**) peraltro, non è stato predisposto alcun modo per ovviare al problema di “causalità” riscontrato già nell’ambito della precedente “VTR 2001-2003”,⁸ secondo cui la valutazione dei referees ha spesso fatto riferimento alla sede di pubblicazione, anziché alla qualità intrinseca dei prodotti. Su questo aspetto si rimanda a oltre, sezione 3 del presente documento.

⁸ http://vtr2006.cineca.it/rel_area/panel_13.pdf

h) (**contraddizioni**) infine, tale decisione appare in evidente contraddizione con i *Criteri* stessi del GEV13, nella misura in cui, forse alla luce del fatto che l'area 13 è considerata non bibliometrica, si enuncia: "I prodotti di ricerca suscettibili di valutazione bibliometrica *non sono automaticamente* (cioè utilizzando in automatico la classe finale suggerita dall'applicazione dell'algoritmo bibliometrico) attribuiti alle classi di merito [...]. L'attribuzione si basa invece sul giudizio esperto del GEV che utilizzerà ogni possibile elemento di valutazione oltre agli indicatori bibliometrici" (p. 9, corsivo aggiunto).

1.2. Equità e non discriminazione nella valutazione

Il GEV 13 ha responsabilità di valutare 19 SSD e 15 SC. Ha però, per ragioni non espresse in alcun documento o comunicazione, deciso di attribuire diverso peso e diversa rilevanza ai vari settori (di nuovo, confondendo la valutazione della ricerca con la definizione della politica culturale, che esula dal mandato del GEV 13 e della stessa ANVUR). Di seguito alcuni dei modi principali in cui ciò è avvenuto.

a) (**discriminazione**) per le riviste italiane, vi è un evidente problema di equità di trattamento. Infatti, nove riviste (di cui due della sotto-area Economia e una di Storia economica) originariamente classificate in fascia D ("accettabile") sono state 'promosse' in fascia C ("discreto"). Questo al fine di raggiungere un numero prefissato nei *Criteri* ("tra 20 e 25", p. 14) di riviste italiane classificate nelle prime tre fasce di valutazione. In realtà c'è solo una rivista italiana nella fascia B ("elevato") in almeno uno dei 4 ranking di economia (New Medit), e solo una nei 4 ranking per storia (Journal of European Economic History); nessuna rivista italiana è in fascia A ("eccellente"); e le altre sono in classe C o inferiore.

A tal proposito, occorre ribadire che la decisione di 'promuovere' le riviste italiane rispetto all'applicazione meccanica di indicatori bibliometrici è a nostro parere corretta, al fine di stabilire un'equità di trattamento a fronte di situazioni oggettivamente diverse, ovvero per evitare una discriminazione indiretta. Questo perché è noto che le riviste in lingua inglese sono sistematicamente premiate da una valutazione fondata sulla visibilità, indipendentemente dalla qualità. Ma l'applicazione selettiva di tale promozione, a sole 9 riviste, determina un'ingiustificata discriminazione diretta: tra le riviste italiane promosse *in quanto* italiane, e le riviste italiane non promosse *nonostante* italiane.

Come mostrato nella tabella 1 per le riviste italiane di economia, l'irrazionalità di tale procedura emerge dalla considerazione che, di fatto, il GEV 13 ha promosso di una classe di merito le riviste italiane che esso stesso considera di 'media qualità': non – per così dire – le 'migliori', che hanno indici citazionali sufficienti a porle automaticamente in fascia C, semplicemente applicando l'algoritmo, ma nemmeno le 'peggiori',⁹ ovvero quelle che sono relativamente più lontane dalla fascia C. Molto meno discriminatorio sarebbe stato, ad esempio, promuovere nella stessa misura (in questo caso, di una fascia) tutte le riviste italiane, indipendentemente dalla loro collocazione iniziale, oppure standardizzare gli indicatori bibliometrici rispetto alla lingua di pubblicazione.

⁹ Dati i criteri descritti prima, di cui ribadiamo di non condividere molti aspetti, a partire dalla confusione tra qualità e visibilità della ricerca. Cionondimeno, è evidente che se alcune riviste sono "valutate" con tali criteri, devono esserlo tutte.

Tabella 1. Riviste italiane di economia in classe C e D (in rosso le "promozioni" di fascia)

Title	Promossa	IF5	AIS	IPP	SJR	h-index	Classe IF5	Classe AIS	Classe IPP	Classe SJR
international review of economics	NO	0.775	0.360	0.420	0.389	16	C	C	C	C
labour	NO	0.713	0.324	0.396	0.265	15	C	C	C	C
psl quarterly review (formerly bnl quarterly review)	NO	0.594	0.256	0.474	0.405	13	C	C	C	C
new medit	parzialmente	0.354	0.079	0.934	0.810	12	C	C	B	B
moneta e credito	NO	0.480	0.195	0.380	0.332	11	C	C	C	C
rivista italiana degli economisti	NO	0.480	0.195	0.380	0.332	11	C	C	C	C
research in economics	parzialmente	0.425	0.166	0.608	0.340	10	C	C	C	C
stato e mercato	SI	0.425	0.166	0.269	0.213	10	C	C	C	C
qa – rivista dell'associazione rossi-doria	SI	0.372	0.140	0.291	0.261	9	C	C	C	C
economia politica	parzialmente	0.485	0.059	0.396	0.236	8	C	C	C	C
Journal of Entrepreneurial and Organizational Diversity	NO	0.320	0.115	0.249	0.227	8	D	D	D	D
terra economicus	NO	0.320	0.115	0.249	0.227	8	D	D	D	D
scienze regionali	NO	0.320	0.115	0.171	0.229	8	D	D	D	D
economic notes	NO	0.270	0.092	0.217	0.164	7	D	D	D	D
rivista di politica economica	NO	0.270	0.092	0.208	0.193	7	D	D	D	D
the european journal of comparative economics	NO	0.270	0.092	0.208	0.193	7	D	D	D	D
economia e politica industriale	NO	0.270	0.092	0.074	0.147	7	D	D	E	D
politica economica	NO	0.270	0.092	0.026	0.108	7	D	D	E	E
economia agro-alimentare	NO	0.221	0.072	0.170	0.161	6	D	D	D	D
giornale degli economisti e annali di economia	NO	0.221	0.072	0.170	0.161	6	D	D	D	D
la rivista delle politiche sociali	NO	0.221	0.072	0.170	0.161	6	D	D	D	D
rivista italiana di politiche pubbliche	NO	0.221	0.072	0.170	0.161	6	D	D	D	D
trasporti	NO	0.221	0.072	0.170	0.161	6	D	D	D	D
ALMATOURISM	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
argomenti	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
economia & lavoro	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
economia e società regionale	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
economia italiana	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
mercato concorrenza regole	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
rivista di diritto ed economia dello sport	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
rivista di economia agraria	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
rivista di economia e statistica del territorio	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
rivista economica del mezzogiorno	NO	0.175	0.053	0.133	0.129	5	D	D	D	D
rivista internazionale di scienze sociali	NO	0.175	0.053	0.085	0.114	5	D	D	E	D

Nota: la promozione 'parziale' riguarda le riviste automaticamente classificate in fascia C per almeno un indicatore, di cui la classificazione per gli altri indicatori è stata portata alla fascia C. Non ha impatto sul ranking finale.

b) **(calibrazione e standardizzazione)** in linea con l'opinione prevalente nella letteratura scientometrica, nel bando della VQR l'ANVUR richiede ai GEV dei settori bibliometrici di standardizzare i singoli valori citazionali rispetto ai diversi ambiti di ricerca presenti in ogni disciplina, e poi di "calibrare" le fasce di merito tenendo conto dei ranking standardizzati sia delle riviste che degli articoli (pp. 15-16). Tale indicazione non si applica però all'area 13, area non bibliometrica, che ha deciso di applicare la bibliometria sua sponte. Questo genera l'assurdo che l'area 13 può applicare la bibliometria con criteri diversi e con maggiore discrezionalità degli stessi settori bibliometrici.

Il GEV 13 ha deciso di non applicare alcuna normalizzazione, nonostante anche per l'economia sia stato mostrato che ambiti diversi di ricerca ricevono sistematicamente un diverso numero di citazioni, indipendentemente dalla qualità dei singoli prodotti di ricerca.¹⁰ A titolo di esempio, nella figura 1 riportiamo la distribuzione per fasce di valutazione delle riviste della sotto-area "Storia economica", distinguendo tra i due suoi principali ambiti: la storia del pensiero economico ("HET", ssd SECS-P/04) e la storia economica ("EH", ssd SECS-P/12).¹¹ Come si vede, per nessuno dei 4 indicatori considerati, nessuna delle riviste di storia del pensiero economico (HET) è classificata in fascia A. Tale sorprendente (e deludente) risultato sarebbe solo interpretabile ipotizzando, in maniera alquanto bizzarra, che l'intera comunità di storici del pensiero economico non produca

¹⁰ Per l'Italia, si veda Corsi M., D'Ippoliti C. e Lucidi F. (2010), "Pluralism at Risk? On the Evaluation of Economic Research in Italy", *American Journal of Economics and Sociology*, vol. 69, n. 5, pp. 1495-1529.

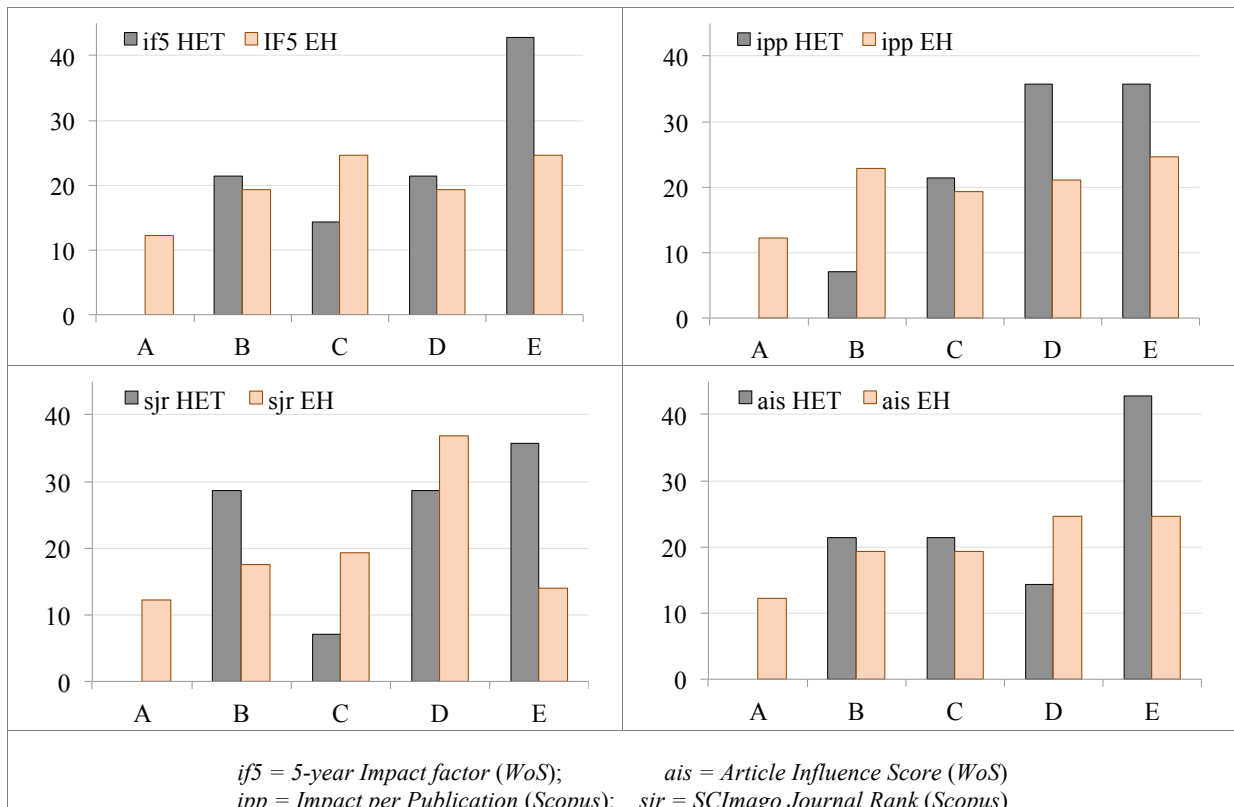
¹¹ La classificazione è stata operata considerando il campo "SU – Subject Descriptor" degli articoli indicizzati in EconLit, considerando una rivista come "EH" o "HET" a seconda della percentuale di articoli pubblicati nei due ambiti nel periodo di riferimento (2011-2014).

praticamente mai (in un orizzonte quadriennale, qual è quello qui considerato) prodotti di qualità “eccellente”, a differenza degli storici economici.¹²

Dunque, è di tutta evidenza che la suddivisione dei ranking delle riviste di area 13 in sole 4 sotto-aree è assolutamente insufficiente (né, d'altra parte, è necessaria una divisione dei ranking, essendo invece sufficiente una standardizzazione degli indicatori). Certamente, nel campo delle scienze economiche i due database considerati suddividono le aree tematiche (le Subject Categories di WoS, e le All Science Journal Classifications di Scopus) in misura insufficiente. Sarebbe però possibile utilizzare i “Subject Descriptors” o alcune aggregazioni dei JEL codes, predisposti nell’ambito di EconLit dall’American Economic Association.¹³ Alternativamente, la classificazione per settori European Research Council (ERC), cui gli stessi *Criteri* del GEV 13 fanno riferimento (p. 6), oppure una tra le più autorevoli classificazioni dei vari ambiti dell’economia: ad es. quella proposta dalla Società Italiana degli Economisti (SIE) nell’ambito del suo ranking delle riviste,¹⁴ o quella utilizzata nel ranking delle riviste del Centre National de la Recherche Scientifique (CNRS) francese.¹⁵

Come già notato, il DM 458 richiede di valutare i “livelli di qualità” dei prodotti di ricerca con riferimento alla “produzione scientifica internazionale” della “comunità scientifica di riferimento” (art. 5). Dunque, la mancanza di una standardizzazione per ambiti di ricerca (come in questo caso la distinzione tra HET e EH) è compatibile solo con l’ipotesi che storici del pensiero economico e storici economici appartengano alla stessa comunità scientifica, ipotesi palesemente falsa.

Figura 1. Valutazione delle riviste della sotto-area “Storia economica”, distinguendo la storia del pensiero economico (HET) dalla storia economica (EH)



¹² Questo ovviamente a parità di probabilità (scarsa per entrambi) di pubblicazione in riviste appartenenti a ranking di altre sotto-aree, e tenendo conto che solo l’1% dei singoli articoli può essere promosso rispetto alla valutazione della rivista.

¹³ <https://www.ebscohost.com/academic/econlit-with-full-text>

¹⁴ <http://www.siecon.org/online/en/commissioni/commissione-per-luniversita-e-la-ricerca/documentazione-interna/>

¹⁵ https://www.gate.cnrs.fr/IMG/pdf/categorisation37_juil_2015-4.pdf

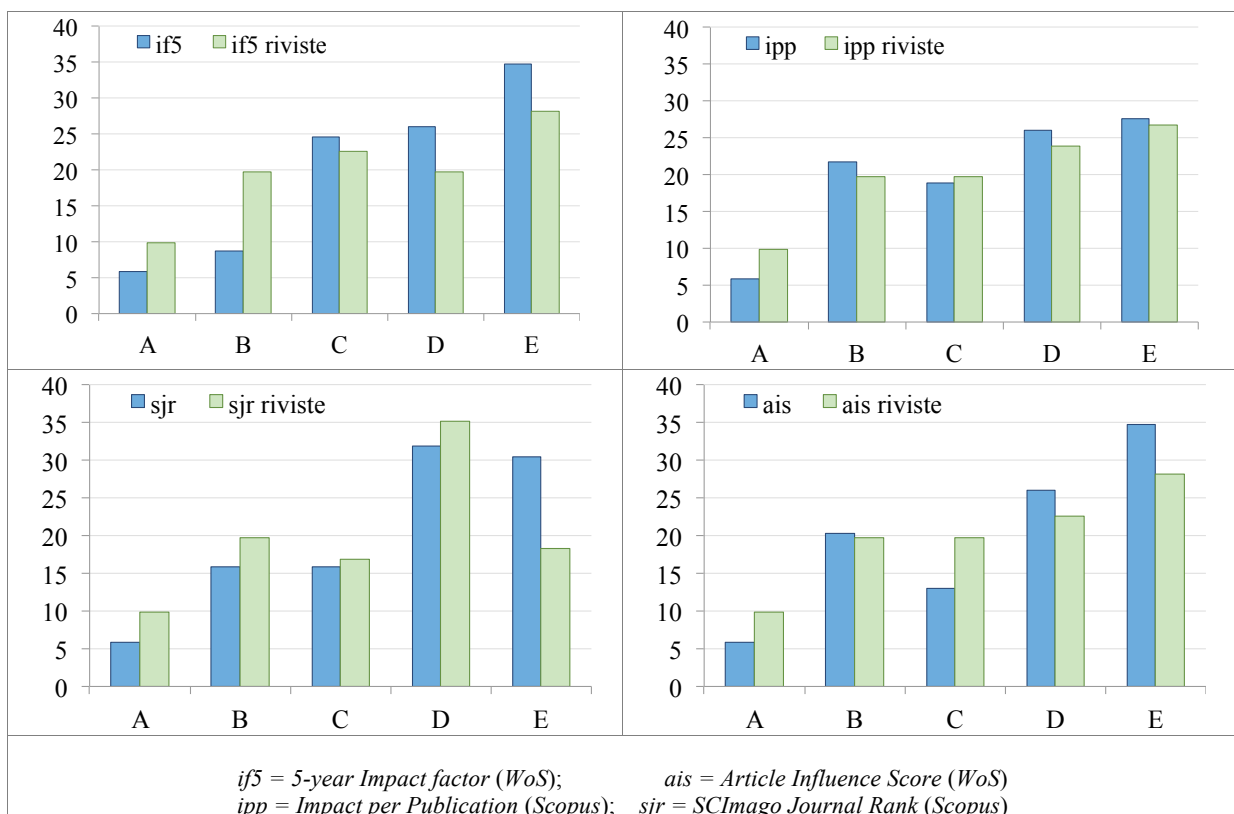
c) **(fasce di merito: articoli vs. riviste)** la “produzione scientifica internazionale” di cui all’art. 5 del DM 458 (citato prima), che dovrebbe costituire il riferimento per la classificazione dei prodotti in fasce di merito, non fa riferimento alle riviste, ma ai singoli prodotti. Anche limitandosi agli articoli su rivista, la differenza è comunque molto rilevante, dal momento che il numero di articoli pubblicati in un dato anno da ogni rivista è molto variabile.

Come si vede nella figura 2, prendendo ad esempio le riviste della sotto-area “Storia economica”, la considerazione dei decili di riviste anziché dei decili di articoli comporta per tutti e 4 gli indicatori citazionali una crescita media della valutazione. E’ da notare che le differenze qui segnalate non riguardano la valutazione del singolo articolo: l’attribuzione in classi nella figura 2 è comunque basata esclusivamente sugli indicatori della rivista; dunque la differenza osservabile nelle due distribuzioni è costituita esclusivamente da un “effetto composizione”. Questo non è che un esempio di come piccole variazioni nei criteri di utilizzo degli indici bibliometrici producano potenzialmente enormi differenze nella valutazione e nei relativi ranking di merito.

Nell’esempio, per tutti gli indicatori la fascia A (prodotti eccellenti) contiene il 5.8% degli articoli considerati per la sotto-area, anziché il 9.8% (che è la proporzione di riviste classificate in fascia A per la sotto-area Storia economica dal GEV13, come approssimazione del 10% richiesto dal DM 458). Parallelamente, di fatto molti più articoli che riviste vengono classificati in fascia E (ovvero di valore “limitato”): tra il 28% e il 35% degli articoli della sotto-area, a seconda dell’indicatore scelto, rispetto al 18%-28% delle riviste. E’ molto probabile, dunque, che ex post la suddivisione per decili delle fasce di merito dei prodotti valutati tramite bibliometria non corrisponda a quella definita ex ante dal decreto ministeriale, e risulti invece inutilmente punitiva.

Questo, peraltro, avrebbe ripercussioni anche sulla valutazione degli enti di ricerca e dei dipartimenti universitari in economia rispetto agli altri (dunque, di nuovo, con conseguenze in termini di politica universitaria).

Figura 2. Confronto delle fasce di valutazione considerando i decili di riviste oppure i decili di articoli, sotto-area “Storia economica”



- d) **(incompletezza della lista)** perfino per i ranking della sotto-area “Storia economica”, nella definizione della lista delle riviste rilevanti indicizzate in WoS, basata sulle Subject Categories (SC) di tale database, non è stata considerata la SC “History of social sciences”. Tale esclusione determina l’assenza nel ranking di alcune riviste che ospitano articoli di storia del pensiero economico (già in numero particolarmente esiguo in WoS, senza bisogno di restringerne ulteriormente il campo).
- e) **(imputazione)** per le riviste non indicizzate in WoS e/o Scopus, la procedura di imputazione degli indici citazionali si basa sull’indice H tratto dal più ampio database Google Scholar (GS). La metodologia utilizzata per tale imputazione contiene però numerosi limiti metodologici che ne inficiano pesantemente i risultati. Anzitutto, molte riviste sono indicizzate in almeno uno tra Scopus e Web of Science. Come mostrato nella tabella 2, la correlazione tra gli indici tratti da WoS e Scopus è molto alta (e spesso maggiore di quella tra gli indici tratti da una delle due fonti e Google Scholar). Invece, sarebbe stato opportuno procedere con una imputazione multipla; aver trascurato questa informazione, decidendo di utilizzare un solo indice su cui basare l’imputazione, potrebbe aver generato una imputazione meno che ottimale.

Tabella 2. Correlazione tra gli indici tratti dai diversi database (lista riviste GEV 13, 22/1/2016)

	IF5 (WoS)	AIS (WoS)	IPP (Scopus)	SJR (Scopus)
IF5 (WoS)	1			
AIS (WoS)	0.8196	1		
IPP (Scopus)	0.9764	0.7768	1	
SJR (Scopus)	0.7189	0.9094	0.7211	1
H-index (GS)	0.8686	0.7162	0.8878	0.672

Ma ben più rilevante, in termini di ranking, è l’aver totalmente ignorato il problema dell’inevitabile incertezza che caratterizza ogni modello statistico: evidentemente, incluso il modello (di regressione) utilizzato per l’imputazione. Infatti, se anziché considerare per ogni rivista il valore puntuale delle prediction dei 4 indicatori, si fossero considerati gli intervalli di confidenza in cui tali valori si trovano con una certa probabilità (ad es. il 95%), si sarebbe immediatamente notato che tali intervalli si sovrappongono per molte riviste attualmente classificate in fasce di merito diverse. Dunque, la procedura attuata (di prendere i valori puntuali) genera la discriminazione diretta tra riviste di cui non si può sostenere, sulle basi dello stesso modello statistico, che abbiano valori diversi, e che cionondimeno sono state classificate in fasce di merito diverse.

- f) **(fonti dei dati)** con l’imputazione di cui sopra, il GEV 13 ha tentato di ovviare al problema della non indicizzazione di un alto numero di riviste in Scopus e/o WoS. Ma il GEV non ha invece ritenuto di considerare il problema della non indicizzazione in Google Scholar. Infatti, molti prodotti rintracciabili tramite il motore di ricerca, sono pubblicati su riviste di cui comunque non è garantito il computo delle citazioni dal servizio Google Scholar Metrics (GSM). Il GEV 13 ha ritenuto di ottenere i dati mancanti relativi agli indici H per queste riviste utilizzando il software “Publish or Perish” (PoP, “Comunicato del GEV13 del 20 novembre 2015”, p. 1). Come mostrato nella tabella 3, l’imputazione dei valori citazionali ha utilizzato un indice H da questa fonte per il 34% dei prodotti (con variazioni notevoli tra le aree).
Però, è ben noto che il software PoP restituisce risultati diversi da GSM. A titolo esemplificativo, proponiamo un confronto tra i valori dell’indice H calcolati con GSM e con PoP per le riviste appartenenti alla sotto-area Storia economica. E’ opportuno sottolineare che la corretta imputazione del dato relativo al valore dell’indice H è particolarmente sensibile, e condiziona fortemente la classificazione delle riviste nelle singole fasce di merito. Ad esempio, per le riviste non indicizzate né in Scopus né in WoS della sotto-area, un indice H pari a 3 implica la classificazione in fascia E (“limitato”), mentre un indice pari a 4 implica la classificazione in fascia D (“accettabile”), così come l’incremento dell’indice H da 6 a 7 implica la promozione dalla fascia D alla C. Come è

evidente dalla tabella 4, PoP tende a sovrastimare il valore dell'indice H. Questo genera, nuovamente, un'ingiustificata discriminazione tra le riviste, sulla base della fonte dei dati utilizzare.

Si sono, infine, verificate delle incongruenze tra l'imputazione da parte del GEV 13 dell'indice H calcolato con PoP e quanto estratto da un'interrogazione dal software PoP. Tale divergenza potrebbe essere dovuta ad es. alla diversa data di interrogazione del software PoP. Alla luce della sensibilità del dato, di cui si è detto, sarebbe opportuno esplicitare la data di estrazione degli indicatori per ogni rivista (criterio esplicitato in GSM, che alla data odierna considera le citazioni al giugno 2015).

Tabella 3. Riviste non indicizzate, per area

		Frequenza	Percentuale
Administration and Management	no WoS	890	73%
	no SCOPUS	456	37%
	né in SCOPUS né in WoS	449	37%
	solo in SCOPUS o solo in WoS	775	63%
	no H-index GS	497	41%
	no H-index GS tra no Scopus/WoS	301	25%
	in % dei no Scopus o WoS	39%	
	Totale	1224	
Economics	no WoS	506	59%
	no SCOPUS	273	32%
	né in SCOPUS né in WoS	270	31%
	solo in SCOPUS o solo in WoS	594	69%
	no H-index GS	312	36%
	no H-index GS tra no Scopus/WoS	181	21%
	in % dei no Scopus o WoS	30%	
	Totale	864	
History	no WoS	46	65%
	no SCOPUS	21	30%
	né in SCOPUS né in WoS	21	30%
	solo in SCOPUS o solo in WoS	50	70%
	no H-index GS	28	39%
	no H-index GS tra no Scopus/WoS	16	23%
	in % dei no Scopus o WoS	32%	
	Totale	71	
Statistics and Mathematical Methods	no WoS	288	51%
	no SCOPUS	158	28%
	né in SCOPUS né in WoS	152	27%
	solo in SCOPUS o solo in WoS	417	73%
	no H-index GS	214	38%
	no H-index GS tra no Scopus/WoS	121	21%
	in % dei no Scopus o WoS	29%	
	Totale	569	
Totale	no WoS	1730	63%
	no SCOPUS	908	33%
	né in SCOPUS né in WoS	892	33%
	solo in SCOPUS o solo in WoS	1836	67%
	no H-index GS	1051	38%
	no H-index GS tra no Scopus/WoS	619	23%
	in % dei no Scopus o WoS	34%	
	Totale	2731	

Tabella 4. Confronto dei valori dell'indice H in Google Scholar, calcolati con Google Scholar Metrics (GSM) e con Publish-or-Perish (PoP), per le riviste della sotto-area Storia economica

Title	issn	GS h-index	PoP h-index	Differenza
journal of economic history	0022-0507	21	25	4
explorations in economic history	0014-4983	20	25	5
economic history review	0013-0117	19	24	5
european review of economic history	1361-4916	15	17	2
historical social research	0172-6404	15	17	2
business history	0007-6791	14	16	2
business history review	0007-6805	13	14	1
journal of economic methodology	1350-178X	12	13	1
JOURNAL OF GLOBAL HISTORY	1740-0228	12	13	1
accounting history	1032-3732	11	12	1
ANNALES. HISTOIRE, SCIENCES SOCIALES	0395-2649	11	10	-1
financial history review	0968-5650	11	11	0
Historical Materialism	1465-4466	11	13	2
journal of management history	1751-1348	11	13	2
MANAGEMENT & ORGANIZATIONAL HISTORY	1744-9359	11	12	1
history of political economy	0018-2702	11	12	1
journal of interdisciplinary history	0022-1953	11	12	1
the european journal of the history of economic	0967-2567	11	11	0
journal of modern italian studies	1354-571X	10	11	1
journal of the history of economic thought	1053-8372	10	12	2
labor history	0023-656X	10	11	1
TECHNOLOGY AND CULTURE	0040-165X	9	10	1
history of economic ideas	1122-8792	8	8	0
enterprise and society	1467-2227	8	8	0
HISTORY COMPASS	1478-0542	7	9	2
entreprises et histoire	1161-2770	7	10	3
investigaciones de historia económica	1698-6989	7	8	1
scandinavian economic history review	0358-5522	7	9	2
rivista di storia economica	0393-3415	6	6	0
INTERNATIONAL LABOR AND WORKING-CLASS	0147-5479	6	7	1
australian economic history review	0004-8992	6	8	2
international review of social history	0020-8590	6	6	0
journal of the economic and social history of the	0022-4995	6	5	-1
journal of tourism history	1755-182X	5	6	1
Indian Economic and Social History Review	0019-4646	5	6	1
labour history	0023-6942	5	6	1
cahiers d'économie politique	0154-8344	4	5	1
Agricultural History Review	0002-1490	4	5	1
Historical Records of Australian Science	0727-3061	4	4	0
international journal of maritime history	0843-8714	3	4	1
textile history	0040-4969	3	4	1
Vierteljahresschrift für Sozial und	0340-8728	2	2	0
società e storia	0391-6987	1	1	0

1.3 Valutazione dell'articolo vs valutazione della rivista

Il DM 458 (linee guida VQR 2011-2014) all'art. 5, enunciando gli “adempimenti dei gruppi di esperti della valutazione”, fa riferimento alla necessità di utilizzare “informazioni bibliometriche [...] relative al prodotto di ricerca [...] e una misura dell'impatto della rivista ospitante il prodotto [...]” (corsivo e sottolineatura aggiunti). E infatti, tutti i GEV bibliometrici utilizzano informazioni sia sulle citazioni individualmente ricevute da ogni singolo prodotto, sia sulle sedi di pubblicazione. Con il comunicato del 29 gennaio 2016, il GEV 13 annuncia di allinearsi a tale pratica tramite rilevazioni delle citazioni al singolo prodotti di ricerca in Scopus e/o WoS. Però, il GEV 13 è l'unico ad aver deliberato di utilizzare le informazioni sui singoli prodotti solo, al massimo, per l'1% degli articoli su rivista indicizzata (“Criteri”, p. 14). A tal proposito, occorre notare:

- a) **(non contraddizione)** ammettendo l'opportunità, almeno per un certo numero di articoli, di utilizzare metriche a livello di prodotto anziché (o in aggiunta a quelle) relative alla sede di pubblicazione, il GEV 13 riconosce l'impossibilità di valutare il singolo articolo sulla base della rilevanza della rivista su cui è pubblicato, aspetto ormai ampiamente acquisito in letteratura.¹⁶

¹⁶ Il tema è molto ampio, e ad es. ha dato origine al movimento dell'*altmetrics*: <http://altmetrics.org/manifesto/> Ad esempio

Ritiene però di correggere tale errore metodologico solo per una percentuale infima di prodotti, di fatto irrilevante ai fini dei ranking complessivi, senza giustificare tale decisione né spiegare il criterio con cui si è prestabilito il limite dell'1%.

- b) **(imprecisioni sulle fonti di dati)** il GEV 13 esclude che la correzione possa riguardare gli articoli pubblicati in riviste non indicizzate in Scopus o Web of Science (WoS). Questo perché “informazioni sulle citazioni individuali sono disponibili solo per le riviste indicizzate” (*Criteri*, p. 14).¹⁷

In realtà, il GEV 13 stesso riconosce la falsità di tale affermazione, nel momento in cui integra il ranking delle riviste indicizzate tramite imputazione dei valori bibliometrici per le riviste non indicizzate, a partire dall'indice H calcolato su Google Scholar. Sembra superfluo rammentare qui che, evidentemente, l'indice H delle riviste è calcolato da Google Scholar Metrics (o Publish or Perish) a partire dalle citazioni ai singoli articoli presenti in Google Scholar. Ma il GEV 13 non spiega in alcun luogo perché Google Scholar possa essere usato per calcolare i valori citazionali aggregati delle riviste, ma non quelli individuali dei prodotti, o perché le citazioni individuali calcolate da Scopus e WoS siano utilizzabili, mentre quelle calcolate da altri database (ad esempio RePEc) no.

- c) **(selezione avversa)** il GEV 13 determina che l'uso dei dati citazionali relativi ai singoli prodotti possa determinare una correzione della valutazione del prodotto (massimo per l'1% dei casi, come si è detto), *solo al rialzo*. Questo espone a un evidente problema di selezione avversa: per cui i numerosi articoli mai citati anche se pubblicati su riviste prestigiose continueranno a essere considerati di buona o eccellente qualità, nonostante, per l'appunto, i dati bibliometrici dicano il contrario.

E' di tutta evidenza che l'unico incentivo per i ricercatori, con questa metodologia, diviene non più svolgere ricerche di buona qualità, anche contro o che superino posizioni consolidate, ma dedicarsi a ricerche *pubblicabili* e *domandate* dal ristretto numero di editors delle riviste attualmente considerate “prestigiose”, e di cui la metodologia attuale esclude per definizione (anzi, per “criterio”) anche solo l'ipotesi di downgrading o di perdita di prestigio. Oltre che dal punto di vista culturale, tale aspetto è problematico anche alla luce della struttura oligopolistica dell'industria editoriale scientifica, il cui grado di concorrenza viene con questo criterio ulteriormente ridotto.

2. ALCUNE CONSEGUENZE DELLA METODOLOGIA ADOTTATA

E' noto che i database Scopus e WoS non rappresentano adeguatamente l'intera produzione scientifica internazionale in campo economico. Per ovviare a tale carenza, il GEV 13 ha deliberato di integrare il proprio ranking delle riviste inserendone diverse tratte da Google Scholar. Per via dei criteri descritti nella sezione precedente, tale integrazione è avvenuta in maniera tale da generare, o non compensare pienamente, gli squilibri tra aree e tra settori-scientifico disciplinari.

Ad esempio, come mostrato nella tabella 5, standardizzando il numero di riviste considerate per ogni sotto-area, per il numero di ricercatori e professori strutturati per ogni SSD (nel 2013 e 2014), emergono differenze notevoli. Così, mentre nella sotto-area statistica sono considerate circa 1,5 riviste per ogni ricercatore, nell'area aziendale e in quella economica solo la metà (circa 0,7), e nell'area storica un terzo di tale valore (appena circa 0,4).

per il caso delle riviste di economia, si veda Oswald A.J. (2007), “An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-Makers”, *Economica*, vol. 74, n. 293, pp. 21-31. Per tutte le altre discipline, si rimanda ai lavori citati in nota 5.

¹⁷ Per una rassegna degli indici per la valutazione dei singoli prodotti di ricerca si veda Wilsdon (2015), *op. cit.* (p. 152).

Tale squilibrio si riflette sul numero di lavori per ricercatore che ex ante potrebbero essere valutati come “eccellenti”, misurato ad esempio dal numero di riviste per docente classificate in fascia A per almeno un indicatore.¹⁸ dal momento che la definizione delle fasce di merito è basata sui decili di distribuzione delle riviste considerate, un campione più grande genera per definizione decili di numerosità maggiore. Come mostrato nella tabella 5, vi sono alcune differenze rispetto ai valori riportati in precedenza, perché la sovrapposizione tra i ranking basati sui 4 indicatori è diversa tra le aree, ma rimane ad esempio il risultato che la sotto-area Storia economica presenta la metà delle riviste in fascia A pro-capite (0,05) rispetto all'economia o all'economia aziendale (0,1). Anche in questo caso, inoltre, per evitare ulteriori effetti distorsivi dovrebbe essere preso in considerazione il numero di articoli pubblicati dalle singole riviste.

Oltre alla disparità di trattamento tra aree e settori scientifico-disciplinari, si pone anche un problema di parità di trattamento tra le riviste. Di fatto, esistono 4 ranking per le riviste indicizzate, basati su due fonti di dati (WoS e Scopus), e un ranking per le riviste non indicizzate, basato sull'indice H di Google Scholar (questo ranking è ottenuto indirettamente, per il tramite dei valori imputati a partire da questo). L'indice H di Google Scholar non è però utilizzato per definire la classificazione in fasce delle riviste indicizzate in Scopus e Thomson Reuters, decisione che genera il risultato che molte riviste sono classificate in una fascia di valutazione più bassa di altre, pur avendo un indice H maggiore. Dunque, quello che è l'unico criterio di merito per le riviste non indicizzate diviene un criterio a volte perfino opposto per le riviste indicizzate.

A titolo di esempio, la tabella 6 riporta le riviste della sotto-area “Economia” classificate in fascia C per tutti e quattro gli indicatori considerati, che hanno un indice H (GS) maggiore di una o più riviste classificate in fascia B per tutti e quattro gli indicatori.

In conclusione, la metodologia adottata dal GEV 13 discrimina tra settori scientifico-disciplinari e tra riviste in maniera difficilmente controllabile ex ante.

Tabella 5. Riviste presenti nel ranking in relazione al numero di ricercatori strutturati, per area e SSD

		2014		2013		Riviste per area		numero riviste per strutturato		riviste in fascia A per strutturato	
		n.	%	n.	%	n.	%	2014	2013	2014	2013
Economics	SECS-P/01	744	17	662	17	864	31.6	0.65	0.73	0.10	0.11
	SECS-P/02	290	6.6	249	6.4						
	SECS-P/03	163	3.7	146	3.8						
	SECS-P/06	130	3	120	3.1						
		1327	30.3	1177	30.3						
History	SECS-P/04	35	0.8	33	0.8	71	2.6	0.37	0.40	0.05	0.06
	SECS-P/12	158	3.6	144	3.7						
		193	4.4	177	4.5						
Administration and Management	SECS-P/07	688	15.8	592	15.2	1224	44.8	0.71	0.82	0.10	0.11
	SECS-P/08	510	11.7	442	11.4						
	SECS-P/09	76	1.7	69	1.8						
	SECS-P/10	139	3.2	108	2.8						
	SECS-P/11	226	5.2	197	5.1						
	SECS-P/13	90	2.1	86	2.2						
	1729	39.7	1494	38.5							
Statistics and Mathematical Methods	SECS-P/05	60	1.4	61	1.6	569	20.8	1.46	1.56	0.08	0.09
	SECS-S/01	387	8.9	361	9.3						
	SECS-S/02	21	0.5	23	0.6						
	SECS-S/03	137	3.1	123	3.2						
	SECS-S/04	62	1.4	57	1.5						
	SECS-S/05	61	1.4	54	1.4						
	SECS-S/06	389	8.9	364	9.4						
	1117	25.6	1043	27							
Totale		4366	100	3891	100	2728	100	0.62	0.70	9	10

¹⁸ Ricordiamo che la valutazione del singolo prodotto può differire da quella della rivista solo nell'1% dei casi.

Tabella 6. Riviste classificate in fascia "C" secondo tutti i 4 criteri, con indice H più alto di almeno una delle riviste classificate in fascia "B" secondo tutti i 4 criteri

Title	WoS	Scopus	h (GS)	if5	ais	ipp	sjr
international review of law and economics	1	1	15	C	C	C	C
journal of economic issues	1	1	15	C	C	C	C
journal of industrial relations	1	1	15	C	C	C	C
Journal of Choice Modelling	0	0	15	C	C	C	C
Cato Journal	0	1	15	C	C	C	C
education economics	0	1	15	C	C	C	C
journal of economic studies	0	1	15	C	C	C	C
labour	0	1	15	C	C	C	C
information technology for development	1	1	15	C	C	C	C
european journal of migration and law	1	1	16	C	C	C	C
review of industrial organization	1	1	16	C	C	C	C
review of network economics	0	1	16	C	C	C	C
capital & class	0	1	16	C	C	C	C
local economy	0	1	16	C	C	C	C
public finance review	0	1	16	C	C	C	C
world patent information	0	1	16	C	C	C	C
international review of economics	0	1	16	C	C	C	C
manchester school	1	1	17	C	C	C	C
Economics-The Open Access Open-Assessment E-Journal	1	1	17	C	C	C	C
canadian public policy	1	1	17	C	C	C	C
International Journal of Sociology and Social Policy	0	1	17	C	C	C	C
journal of education and work	0	1	17	C	C	C	C
transfer: european review of labour and research	0	1	17	C	C	C	C
contemporary economic policy	1	1	18	C	C	C	C
international labour review	1	1	18	C	C	C	C
journal of competition law and economics	1	1	18	C	C	C	C
international journal of cultural policy	1	1	18	C	C	C	C
tourism economics	1	1	19	C	C	C	C
china & world economy	1	1	20	C	C	C	C
applied economics	1	1	35	C	C	C	C

3. LA VALUTAZIONE TRAMITE PEER REVIEW

Curiosamente, il GEV 13 non prevede l'utilizzo di metodi bibliometrici per la valutazione di libri o capitoli di libro, nonostante l'esistenza ormai diffusa di ranking di case editrici, e la crescente indicizzazione delle monografie scientifiche e dei libri collettanei negli stessi Scopus e Web of Science. Auspicabilmente, la limitazione della peer review a questi prodotti non è l'anticamera per il loro presente o futuro 'declassamento' a prodotti 'meno valutabili' o 'meno rilevanti'.

Infatti, da un lato nel campo dell'economia i libri rimangono un fondamentale strumento di comunicazione scientifica,¹⁹ e dall'altro la peer review rimane il metodo di valutazione utilizzato (per la valutazione degli articoli) dalle stesse riviste incluse nel ranking del GEV 13 e quello utilizzato per la decisione sull'indicizzazione delle riviste in Scopus e/o WoS. Sarebbe dunque del tutto incoerente ritenere meno valido un metodo di valutazione che è alla base proprio dei metodi (bibliometrici) ritenuti invece validi.

Sui criteri e metodi della valutazione tramite peer review, il documento *Criteri* del GEV 13 è breve e poco dettagliato. Allo scopo, si segnalano le seguenti criticità:

¹⁹ Si pensi ad es. a Piketty T. (2014), *Capital in the XXI Century*, Harvard University Press, che in un anno e mezzo ha già ottenuto 377 citazioni su Google Scholar; o Mazzucato M. (2013), *The Entrepreneurial State*, Anthem Press, che in due anni e mezzo ha ottenuto 426 citazioni sullo stesso database.

- a) (**scarsa trasparenza**) come minimo, la garanzia degli interessi legittimi di tutti gli interessati impone la pubblicazione immediata della scheda che i referee saranno chiamati a compilare, e delle linee guida fornite dal GEV 13 ai referee. Ex post, sarà necessario pubblicare non solo l'elenco complessivo dei referee chiamati a valutare i prodotti, come attualmente previsto dai *Criteri*, ma anche l'indicazione del numero di prodotti valutati da ogni referee. Inoltre, nel rispetto della riservatezza delle informazioni sensibili, ogni autore deve poter conoscere i giudizi dei referee sui propri prodotti, e non solo la valutazione finale del GEV priva di dettagli e note.
- b) (**distinguere qualità e visibilità**) nell'ambito della VTR 2001-2004, richiamata all'inizio, relativamente alla valutazione in peer review è stato denunciato che: (i) i referee hanno confuso la valutazione dell'internazionalizzazione delle riviste con la valutazione della qualità degli articoli; (ii) i referee hanno spesso giudicato la qualità degli articoli principalmente o esclusivamente in base alla loro valutazione discrezionale del prestigio della sede editoriale (*Rapporto Finale per l'area 13*, Appendice n. 4, "A note on points of dissent"). E' dunque necessario che il GEV predisponga delle schede di valutazione appropriate, adeguate ad evitare l'insorgere di tali problemi che violano non solo lo spirito, ma anche la lettera del D.M. 458. Inoltre, statistiche descrittive e network analysis dei membri del GEV e dei valutatori esterni coinvolti negli esercizi passati²⁰ hanno evidenziato una rilevante mancanza di *fairness*, nel senso del coinvolgimento di referee e membri del GEV che abbiano caratteristiche e che rappresentino metodi, approcci e conoscenze sufficientemente diversificati. Inevitabilmente, la selezione dei referee rappresenta un aspetto cruciale ai fini della correttezza e dell'equilibrio dell'intera procedura di valutazione in peer review.

4. POSSIBILI RIMEDI

Anche alla luce delle proposte contenute nel documento redatto da tutte le società scientifiche della sotto-area A (economia aziendale),²¹ proponiamo le seguenti misure minime, per attenuare almeno alcuni degli effetti più dannosi delle distorsioni identificate.

- 1) **garantire la valutazione in peer review** per qualsiasi prodotto ne sia fatta richiesta (e quindi permetterne la richiesta per ogni prodotto). L'area A del GEV 13 è interamente composta da settori non bibliometrici, e non è possibile procedere a valutazioni automatiche sulla base dei ranking di riviste (procedura peraltro rinnegata dal GEV 13 stesso). Inoltre, tale valutazione deve considerarsi addizionale, e non necessariamente sostitutiva delle informazioni tratte da indici bibliometrici (che possono rimanere un ausilio alla valutazione complessiva).
- 2) **standardizzare per ambito disciplinare tutte le misure citazionali** e gli indici bibliometrici, nonché le future valutazioni in peer review. Questo significa anche rivedere la definizione delle fasce di merito, correggendo i difetti in senso restrittivo dovuti al campione di riferimento internazionale e alle distorsioni segnalate, nonché selezionare con equilibrio i referee, al fine di garantire un'equa rappresentazione di tutti i metodi e approcci senza discriminazioni.

²⁰ Si veda Baccini A. (2014), "La VQR di area13: una riflessione di sintesi", *Statistica & Società*, anno III, n. 3, pp. 32-37.

²¹ "Osservazioni delle società scientifiche area economico-aziendale sui criteri valutativi elaborati dal GEV 13", 25 gennaio 2016; disponibile alla URL <http://www.accademiaidea.it/wpaidea/wp-content/uploads/2016/01/OSSERVAZIONI-25-GENNAIO-Macrosettore-13B.pdf>

- 3) utilizzare **un mix di indicatori bibliometrici di articoli e di riviste** (come fatto da tutti i settori bibliometrici) per tutti i prodotti valutati con la bibliometria, non solo l'1%, e chiarire la fonte e il metodo di reperimento dei dati. E' contraddittorio che, se il metodo di valutazione è la bibliometria (quindi le citazioni), articoli con zero citazioni siano giudicati come "buoni". Allo stesso modo, le fasce di merito vanno definite a livello di articolo (calibrato), e non di rivista, o almeno non solo di rivista.
- 4) applicare **gli stessi criteri e metodi** per tutte le riviste e tutti i prodotti, **senza discriminazioni**, al contrario di quanto è accaduto, ad esempio, per le riviste italiane e per l'uso di indici H da diverse fonti.
- 5) nella peer review, prevedere procedure e metodi atti a prevenire e limitare il più possibile i fenomeni di "causalità inversa", **evitando che i referee valutino gli articoli sulla base della sede di pubblicazione** (così ripristinando indirettamente l'unico criterio bibliometrico). Inoltre, in caso di valutazione sia in peer review che bibliometrica, non è ipotizzabile che prevalga in maniera automatica il secondo criterio. Prevedere le massime forme di pubblicità e trasparenza della procedura.