

Le politiche editoriali delle riviste scientifiche: come valutarle

Test statistici di un modello dell'Impact Factor

di Mario de Marchi e Maurizio Rocchi

Le speranze di applicare delle policy pubbliche che regolino più o meno efficacemente il progresso scientifico e tecnologico si indirizzano sull'efficacia di due aspetti: la capacità di prevedere l'orientamento della scienza e della tecnologia, così da intervenire in anticipo per assecondarne lo sviluppo, e la capacità di discriminare i progetti di ricerca in cui l'uso delle risorse è stato più efficiente, così da selezionare per l'esecuzione di futuri progetti i ricercatori che si sono dimostrati più abili.

Sul primo punto un po' poco sembra possibile alla luce delle attuali conoscenze, che sottolineano la difficoltà estrema di predire il contenuto della scoperta scientifica e dell'innovazione tecnologica (Pavitt).

Sul secondo punto, invece, degli

avanzamenti sia in astratto (Albert et al. Vinkler) sia nell'applicazione pratica si stanno verificando e stanno interessando una parte non marginale della comunità scientifica. Si vanno diffondendo ormai, anche in Italia (De Marchi et al.), processi formalizzati di valutazione *ex post* ed *ex ante* sempre più attenti.

È evidente che una valutazione la più scrupolosa possibile dei frutti della ricerca (nel contesto preso qui in considerazione, le pubblicazioni) andrebbe fatta con un esame qualitativo e quindi approfondito dei singoli prodotti della ReS, il che richiederebbe tempo e competenze di cui non tutti i policy maker solitamente dispongono.

L'informazione che in pratica viene fornita ai decisori deve quindi essere sintetica rispetto a quella che sarebbe ottimale in linea di principio.

A prima vista pare che i decisori potrebbero procurarsi quest'informazione sintetica anche grazie a procedure di analisi qualitativa, delegando cioè il compito di leggere e valutare la massa di informazione scientifica disponibile per la valutazione dei prodotti della ReS a

persone nominate appositamente perché ritenute competenti ("valutatori").

Si può eccepire che questa può essere considerata una procedura in cui le decisioni dei policy maker vengono condizionate dai giudizi soggettivi dei valutatori: è possibile che questa eccezione venga ritenuta irrilevante purché la loro selezione sia accurata.

Ma il dubbio fondamentale viene suscitato proprio dal fatto che per attribuire la competenza di valutatori ad alcuni scienziati (a preferenza di altri) i decisori necessiterebbero delle sintesi qualitative sui titoli accumulati da ciascun candidato a divenire valutatore che solo dei valutatori già riconosciuti come tali sarebbero in grado di preparare, il che comporta un ragionamento circolare.

Delle procedure quantitative si spera proprio che possano sormontare queste difficoltà, con la loro apparente oggettività e facilità di applicazione (Garfield et al., Glanzel, Narin Moed 2).

Il diffondersi della valutazione quantitativa della ricerca è cioè una risposta adattativa dei policy maker alla necessità di aumentare i controlli centralizzati sull'efficienza della spesa pubblica in ReS, non potendo evidentemente la determinazione dell'investimento pubblico basarsi in prima istanza sui meccanismi equilibratori spontanei e decentralizzati del mercato.

È possibile che a questa risposta adattativa dei policy maker possa seguire una contromossa da parte di una frazione della comunità scientifica e di coloro che aspirano a farne parte: la tendenza a sostituire la elevata qualità tipica di un'attività elitaria, quale la ricerca, con la quantità.

Per evitare che questo snaturarsi sia pure in minima parte la ricerca e possa in qualche misura vanificare lo sforzo fatto per valutarla, occorre che i policy maker predispongano

Questo articolo è stato pubblicato con il titolo originale *The editorial policies of scientific journals: testing an impact factor model* sulla rivista "Scientometric", 51 (2001), 2, p. 395-404. Gli autori ringraziano sentitamente Enzo Casolino, direttore della biblioteca del Cnr, per aver messo a disposizione la sua struttura e il personale della biblioteca per la preziosa collaborazione.

no ulteriori controlli che bilancino la controrisposta adattativa alla valutazione data da scienziati poco coscienziosi con la produzione da parte di ricercatori mediocri di grandi numeri di pubblicazioni corrispondenti in realtà a masse crescenti di letteratura praticamente senza valore scientifico.

Per non ricadere nei problemi della valutazione qualitativa già discussi e da cui si intende uscire con la valutazione quantitativa, anche questi ulteriori controlli dovrebbero essere sintetici e di facile applicazione, cioè quantitativi.

Un passo essenziale nella prevenzione delle distorsioni che più facilmente si possono associare alla valutazione quantitativa è l'adozione di procedure quantitative di valutazione delle politiche editoriali delle riviste scientifiche, o meglio delle riviste che si dichiarano tali.

Queste procedure debbono essere concepite in modo da mettere in luce le eventuali distorsioni causate da politiche editoriali delle riviste non ottimali. Con la valutazione quantitativa le riviste, infatti, vengono ad assumere un ruolo cruciale nella formazione delle politiche scientifiche pubbliche e devono quindi essere sottoposte a specifiche misure di monitoraggio che ne controllino l'adeguatezza del comportamento ai requisiti necessari per svolgere l'attività semiufficiale di certificatori della qualità dei prodotti della ricerca.

Come è noto, nel corso di puro e semplice monitoraggio o di valutazioni che presiedono alle scelte di policy, per monitorare l'importanza di ciascuna pubblicazione in modo più aderente alle effettive ricadute che quella pubblicazione avrà nella comunità scientifica, essa viene pesata con un indice che riflette il successo della rivista nel mercato delle pubblicazioni scientifiche, il cosiddetto Impact Factor.

L'Impact Factor (Abt, Balaban, Moed et al., Ramirez et al.) è il pri-

mo e il più noto fra gli indicatori che pur essendo rigorosamente quantitativi hanno lo scopo di misurare la elusiva grandezza "qualità" di un prodotto scientifico.

La complessità dei fenomeni della produzione, della divulgazione e della fruizione della scoperta scientifica implica naturalmente che per riflettere i caratteri molteplici di questo fenomeno ci si può fondare

più che su un unico indicatore, per quanto utile, su una costellazione di indicatori. Per l'interdipendenza di questi fenomeni, c'è da attendersi che gli aspetti rilevati da un ventaglio di indicatori non siano del tutto scollegati fra loro, ma anzi che siano connessi da relazioni funzionali di causa ed effetto.

Nella nostra ricerca abbiamo identificato alcune circostanze che sono correlate all'Impact Factor e che possono essere misurate quantitativamente, in modo da poter figurare in un'ipotesi di determinazione, per quanto parziale e imperfetta, di questo indicatore.

Il nostro scopo nello svolgimento dell'indagine qui presentata era di controllare in quale misura la grandezza più frequentemente usata per pesare l'importanza di una pubblicazione, cioè l'Impact Factor, rifletta effettivamente la qualità di ciascuna rivista.

Nella scelta di queste circostanze siamo stati indirizzati da un'esigenza che ha strettamente a che fare con l'uso di tale indicatore come strumento di policy. Proprio perché l'Impact Factor è venuto assu-



mendo una rilevanza così alta nelle politiche della ricerca di molti paesi (inclusa sempre più l'Italia) è necessario assicurarsi che nella sua distribuzione fra le varie riviste esso rispecchi più o meno fedelmente certe caratteristiche considerate desiderabili dal policy maker. E affinché un alto Impact Factor sia anche senz'altro un indice di successo scientifico, può essere auspicabile che esso si associ ad almeno altri tre caratteri del lavoro scientifico a cui si riferisce.

In primo luogo, l'importanza scientifica del lavoro è chiaramente accresciuta dalla novità dei risultati che in esso sono presentati. In secondo luogo, la rilevanza scientifica implica tendenzialmente una maggior complessità di esposizione e dunque un carattere maggiormente specialistico del lavoro. In terzo luogo, l'importanza scientifica si riflette senz'altro nella capacità del lavoro di imporsi, grazie all'originalità e al rigore del suo contenuto, all'attenzione dei revisori di una rivista scientifica e di ottenere la loro approvazione.

Abbiamo deciso di misurare ➤

questi tre caratteri con tre indicatori strettamente numerici, in accordo con le direttive discusse sopra sull'opportunità di una valutazione quantitativa.

La novità del contributo scientifico si riflette nel primo indicatore: il ritardo medio nella pubblicazione di un lavoro (espresso in settimane) dopo che è stato accettato dai revisori di una rivista. Tanto maggiore è il tempo che intercorre fra i due eventi, tanto minore è la novità dei contenuti, e non solo per il semplice trascorrere del tempo ma anche perché il ritardo segnala il giudizio della rivista stessa nel valutare il contributo innovativo che i suoi articoli danno al progresso della scienza.

Il secondo indicatore ha a che fare con la complessità dei temi di ricerca e con la conseguente complessità e specializzazione del linguaggio necessario a descriverli: si tratta del numero di discipline coperte dalla rivista considerata di volta in volta. Possiamo sensatamente ipotizzare che i contributi di maggior rilevanza scientifica tenderanno a essere pubblicati su riviste con un basso numero di aree disciplinari, e quindi con un pubblico altamente specializzato in grado di apprezzare il contenuto degli articoli nonostante le difficoltà di linguaggio coperte (esistono delle eccezioni, discusse oltre).

Il terzo indicatore misura la difficoltà di ottenere la pubblicazione di un lavoro su una certa rivista in base alla percentuale media di rigetto degli articoli proposti alla rivista: è chiaro che tanto più selettivo sarà il lavoro dei revisori della rivista, tanto maggiore risulterà la qualità media dei contributi che essa pubblica.

La *proxy* da noi impiegata per misurare la variabile dipendente "Impact Factor", infine, era il numero di copie a pagamento diffuse da ciascuna rivista.

Questi tre indicatori sono quindi

entrati a far parte di un modello probabilistico della determinazione dell'Impact Factor; quest'ultimo era cioè la variabile dipendente del modello (ossia l'ipotizzato "effetto") mentre i tre indicatori da noi costruiti erano le variabili indipendenti (e quindi rappresentavano secondo la nostra ipotesi alcune delle "cause" per cui l'Impact Factor assume i valori che assume).

Secondo le ipotesi già discusse, abbiamo ottenuto delle predizioni per il segno dei parametri che dovrebbero spettare alle tre variabili. Il segno spettante al coefficiente dell'indicatore del ritardo nella pubblicazione dei lavori si supponeva essere negativo. Questo segnalava una relazione inversa fra ritardo nella pubblicazione e Impact Factor, poiché, come abbiamo sostenuto sopra, un grande ritardo implicherebbe una scarsa novità e questa a sua volta comporterebbe un limitato interesse della comunità scientifica per la rivista, con un corrispondentemente basso Impact Factor.

Il segno ipotizzato per il coefficiente della variabile "numero di discipline" è negativo. Si supponeva cioè che la relazione fra specializzazione del linguaggio e rilevanza del risultato fosse diretta. Questa è una delle ipotesi più delicate nel nostro modello. Vi sono infatti almeno due casi esemplari di riviste scientifiche ben conosciute con un altissimo Impact Factor e relativamente despecializzate nei contenuti. Ma abbiamo creduto che queste fossero delle eccezioni (da trattare eventualmente come outlier) nel campo dell'editoria scientifica e che esse non invalidassero la relazione probabilistica generale postulata fra numero di discipline coperte e Impact Factor.

Per il parametro della variabile "tasso di rigetto" ci aspettavamo un segno senz'altro positivo, ovvero una relazione diretta fra la qualità

attribuita dal mercato dei fruitori scientifici alla rivista e la sua selettività.

Non ci attendevamo che il modello assorbisse (o come si usa dire nel gergo degli statistici "spiegasse") un'elevata quota della variabilità (in termini statistici più rigorosi "varianza") del fenomeno Impact Factor. Il motivo fondamentale di questa aspettativa sta nella fortissima dispersione della dimensione delle singole comunità scientifiche a cui si riferiscono le riviste, e nel conseguente inevitabile e fisiologico alto valore del residuo casuale della variabile dipendente. Ci saremmo quindi ritenuti soddisfatti, per quel che riguarda il potere predittivo del modello, anche con una quota di variabilità assorbita bassa, purché, naturalmente, statisticamente significativa.

Il campione statistico su cui abbiamo eseguito i test statistici era originato dall'universo delle riviste scientifiche acquistate dal Consiglio nazionale delle ricerche (Cnr) d'Italia. Grazie al fatto che il Cnr è un'istituzione di ricerca multidisciplinare molto articolata, si tratta di un folto gruppo di giornali che rappresentano le pubblicazioni più importanti di praticamente tutte le aree della ricerca logico-matematica, naturale, sociale e umanistica. Il totale delle riviste che componevano l'universo di partenza era di circa 1.800.

Di queste, abbiamo potuto rintracciare un indirizzo che permettesse una rapida esecuzione dell'indagine (fax o e-mail) per 965: in particolare 143 riviste sono state raggiunte via fax e 822 con messaggi di posta elettronica. Dopo il primo invio, dato il ridotto numero di risposte, abbiamo eseguito con gli stessi mezzi ben tre solleciti.

Il questionario inviato, in cui si garantiva ai rispondenti l'assoluta riservatezza a livello di singolo giornale delle informazioni eventualmente comunicate, era composto

Tab. 1 - Domande del questionario somministrato alle riviste scientifiche

1. How much time (in weeks) passes on average between acceptance of an article and effective publication?
2. What percentage of articles are rejected out of all those proposed for publication?
3. How many copies has your journal sold on average over the last year and what has been its turnover in dollars?
4. Which of the following disciplinary areas does the journal cover?
Agriculture; Biotechnology; Computer science; Economics; Engineering; Environmental research; Geology; Humanities Archaeology and history of arts; History; Geography and ethnology; Psychology and Pedagogy; Law; Mathematics; Medicine and Biology; Physical Sciences; Physics and Astronomy; Chemistry; Other (specificate)

dalle quattro domande riportate nella tabella 1.

In totale le risposte complessivamente ottenute sono state 92, con un tasso di risposta del 9,5 per cento. Di queste risposte, 15 provenivano da riviste che sostenevano di non poter divulgare i dati da noi richiesti per motivi di riservatezza; 5 risposte concernevano riviste che pubblicavano esclusivamente abstract di articoli già pubblicati altrove – naturalmente questi *secondary journals* sono stati esclusi (in genere per ammissione delle stesse riviste rispondenti) dal campione definitivo che abbiamo utilizzato per i nostri test statistici. Questo campione era costituito in totale dalle 72 riviste che avevano effettivamente dato risposte utili al questionario somministrato. Il rapporto fra questo campione definitivo e l'universo delle imprese editoriali a cui era stato somministrato il questionario, bisogna ammetterlo, è decisamente basso. Ma ciò, più che sollevare dubbi sulla efficacia della nostra indagine (nella quale, ricordiamo, abbiamo eseguito ben tre solleciti), ci sembra porre delle questioni delicate sul mondo delle riviste scientifiche, le quali non si sono limitate a non collaborare all'indagine semplicemente non rispondendo, ma in un numero significativo di casi si sono prese perfino il disturbo di comunicare che i dati richiesti non potevano essere rivelati perché "confidenziali". Ci chiediamo, e proponiamo la nostra domanda alla riflessione del

lettore: è giustificata tanta riservatezza nelle riviste scientifiche, che, avendo ormai un ruolo semiufficiale nella catena di giudizi e valutazioni che presiedono alla formulazione delle politiche della ricerca, dovrebbero semmai essere dei modelli di nitidezza?

La tabella 2 fornisce i valori medi delle quattro variabili rilevate. Prima ancora di effettuare i test, da una semplice ricognizione sommaria dei valori delle variabili abbiamo riscontrato la plausibile insorgenza di fenomeni di collinearità per la variabile "settimane" e, in misura anche superiore, per la variabile "discipline". Ciò dipende naturalmente dall'addensarsi di molti valori della prima attorno a un ammontare di 16-20 (il che riflette la cadenza trimestrale nella pubblicazione di molte riviste) e di molti valori della seconda su un ammontare pari esattamente a uno (per la tendenziale specializzazione dei giornali scientifici).

I risultati dei test statistici (raccolti

per comodità del lettore nella tabella 3) rispondono parzialmente alle nostre aspettative. La varianza della variabile Impact Factor assorbita dal modello è pari a circa il 10 per cento del totale. È questo un valore certamente non alto, come ci attendevamo; ma, come speravamo, esso risulta significativo dal punto di vista statistico a un livello di significatività poco superiore a quello del 5 per cento. Il modello presenta quindi una certa efficacia come ipotesi di spiegazione di un carattere dall'elevata dispersione quale è l'Impact Factor.

Più complessa, come del resto c'era da attendersi ed era stato preannunciato sopra, è l'interpretazione dei test riguardanti almeno due delle tre variabili esplicative.

Sia per il parametro della variabile "settimane" che ancor più per quello della variabile "discipline", l'addensamento delle risposte all'indagine attorno a certi particolari valori induce, come già preannunciato, problemi di collinearità e quindi di bassa significatività delle stime.

Così, entrambe le stime danno valori negativi, in accordo con le nostre predizioni, ma la probabilità che si abbiano valori distanti dallo zero almeno quanto quelli effettivamente ottenuti con le regressioni (un indice inversamente proporzionale alla cosiddetta "significatività" delle stime) è pari a circa il 21 per cento per il parametro di "setti- ➤

Tab. 2 - Valori medi delle quattro variabili

Discipline	1.86	
Settimane	19.6	
Tasso di rigetto		45.8
Copie vendute complessivamente	9440	

Tab. 3 - Risultati della regressione multivariata

	Stime	P-values
R ² corretto	0.1	0.055
Discipline	-893	0.62
Settimane	-190	0.21
Tasso di rigetto	250	0.04

mane” e a ben il 62 per cento per quello di “discipline”: le due variabili, e in particolare la seconda, sono quindi relativamente poco significative (è cioè relativamente alto il rischio di attribuire loro un effetto sull’Impact Factor che potrebbe essere dovuto al mero caso). Come abbiamo anticipato, risultati del genere sembrano fisiologici con queste particolari variabili e possiamo considerarci moderatamente soddisfatti delle prestazioni del nostro modello a questo riguardo; sembra infatti plausibile che anche con campioni più vasti dell’universo delle riviste scientifiche si riproporrebbero fatalmente i problemi di collinearità già discussi.

I risultati sono senz’altro soddisfacenti per quanto riguarda la stima del parametro della variabile “tasso di rigetto”, che è, come ci aspettavamo, positiva e significativa a un livello di confidenza del 4 per cento. Questa sembra una prova piuttosto robusta del fatto che la selettività dei revisori e la qualità degli articoli pubblicati hanno un rilevante effetto positivo sull’Impact Factor.

Il nostro è stato soltanto un primo tentativo di un controllo empirico che, auspicabilmente, andrebbe replicato più volte, per arrivare sia a un affinamento della metodologia sia a una migliore affidabilità dei risultati. Nell’esprimere questo giudizio ci rifacciamo assai più alle tradizioni delle scienze naturali che a quella delle scienze sociali, dove purtroppo, ancora troppo spesso i test empirici sono degli *unica* irripetuti e quindi poco significativi.

Pur con i suoi limiti di primo esperimento del genere, dal nostro lavoro sembra si possano trarre alcune conclusioni di una certa rilevanza.

In primo luogo, è sconcertante la reazione di molte delle riviste da noi contattate alla richiesta di informazioni, che pure era stata fatta con garanzia dell’anonimato. Parecchie riviste non si sono limi-

tate a evitare di rispondere per semplice trascuratezza, in qualche misura comprensibile vista la gran mole di indagini statistiche e di questionari che circolano in tutte le imprese e il tempo che la loro compilazione richiede: esse invece ci hanno risposto affermando esplicitamente che i dati da noi richiesti erano “confidenziali” e non potevano essere rivelati. Questo eccesso di riservatezza è quanto meno sorprendente in un anello cruciale della catena di attori sul cui operato si fondano e sempre più si fonderanno le decisioni di politica della ricerca prese dall’operatore pubblico. Ci saremmo aspettati, semmai, una grande trasparenza delle riviste nel mettere a nudo le caratteristiche delle loro strategie editoriali. Questa prima conclusione contiene un ammonimento ai policy maker a prestare attenzione ai delicati e complessi meccanismi di funzionamento del mondo delle riviste scientifiche.

In secondo luogo, non sembrano essere emerse, per quel che i dati ci consentono di desumere, strategie distorsive dei meccanismi spontanei della valutazione effettuata dal mercato delle pubblicazioni scientifiche sulla qualità dei lavori. Questo è confortante, dato il rischio che alcuni attori di questo scenario, eventualmente dotati di pochi scrupoli, possano adottare degli stratagemmi per eludere i vincoli posti alla qualità da una rigorosa applicazione di criteri quantitativi nella valutazione della ricerca.

In terzo luogo, più specificamente, abbiamo avuto una conferma dell’affidabilità dell’Impact Factor quale strumento di monitoraggio della qualità della ricerca e come mezzo per valutare la ricerca stessa. Questo appare un risultato rassicurante per l’impostazione corrente delle politiche pubbliche e conferma la sostanziale onestà dei meccanismi con cui si attua la

competizione nell’intrapresa scientifica.

Riferimenti bibliografici

- ABT H., *Do important papers produce high citation count?*, “Scientometrics”, 48 (2000), p. 65-70.
- ALBERT M. et al., *Direct validation of citation counts as indicators of industrially important patents*, “Research Policy”, 20 (1991), p. 251-259.
- BALABAN A., *How should citations to articles in high- and low-impact journals be evaluated, or what is a citation worth?*, “Scientometrics”, 37 (1996), p. 495-498.
- GARLFIELD E. – WELLJAMS-DOROF A., *Citation data: their use as quantitative indicators for science and technology evaluation and policy-making*, “Science and Public Policy”, 19 (1992), p. 321-327.
- GLANZEL W., *The needs for standards in bibliometric research and technology*, “Scientometrics”, 35 (1996), p. 167-176.
- MOED H. et al., *A critical analysis of the journal Impact Factors of Angewandte Chemie and The Journal of the American Chemical Society. Inaccuracies in published Impact Factors based on overall citations only*, “Scientometrics”, 37 (1996), p. 105-116.
- MOED H. et al., *The use of bibliometric data for the measurement of university research performance*, “Research Policy”, 14 (1985), p. 131-149.
- NARIN F., *Technology indicators in strategic planning*, “Science and Public Policy”, 19 (1992), p. 369-382.
- PAVITT K., *The inevitable limits of EU R&D funding*, “Research Policy”, 27 (1998), p. 559-568.
- RAMIREZ A. – GARCIA E. – DEL RIO A., *Renormalized Impact Factor*, “Scientometrics”, 47 (2000), p. 3-9.
- VINKLER P., *Model for quantitative selection of relative scientometric impact indicators*, “Scientometrics”, 36 (1996), p. 223-236.