Appendice B. Il confronto tra valutazione *peer* e valutazione bibliometrica

I GEV che hanno utilizzato gli indicatori bibliometrici per la valutazione degli articoli indicizzati in ISI WoS e Scopus hanno selezionato un campione casuale di dimensione pari a circa il 10% degli articoli valutati bibliometricamente e li hanno sottoposti alla valutazione *peer*. L'obiettivo è il confronto tra le due metodologie di valutazione applicate allo stesso campione di articoli, allo scopo di valutare il grado di corrispondenza dei risultati. Nel seguito, saranno presentati i risultati in forma aggregata e distinti per area.

B.1 Il campionamento statistico

Un campione casuale di articoli su rivista sottoposti a valutazione bibliometrica è stato estratto dalla popolazione di 77.159 articoli, valutabili bibliometricamente e sottoposti alla valutazione nei GEV che hanno utilizzato indicatori bibliometrici. Il campione è stato stratificato in base alla distribuzione dei prodotti all'interno dei sub-GEV individuati nelle varie aree. Ai fini della stratificazione, gli articoli sono stati attribuiti ai sub-GEV sulla base del settore scientifico-disciplinare (SSD) nel quale sono stati valutati. Complessivamente, non tutti i prodotti del campione teorico sono stati referati entro i termini previsti dall'esercizio di confronto tra le due valutazioni. Il campione empirico osservato è quindi risultato essere pari al 9,3% degli articoli sottoposti a valutazione bibliometrica nelle aree "bibliometriche". La Tabella B.1 riporta l'elenco dei GEV bibliometrici e, per ciascuno di essi, la dimensione della popolazione, del campione empirico effettivamente referato in valori assoluti, percentuali e in percentuale sulla popolazione per il campione empirico.

GEV	Popolazione	% nella popolazione	Campione empirico	% nel campione empirico	% del campione empirico sulla popolazione
1	4.631	6,0	444	6,2	9,6
2	10.182	13,2	1.008	14,1	9,9
3	6.625	8,6	653	9,1	9,9
4	3.953	5,1	388	5,4	9,8
5	10.423	13,5	951	13,3	9,1
6	15.400	20,0	1.293	18,0	8,4
7	6.354	8,2	630	8,8	9,9
8b	2.370	3,1	234	3,3	9,9
9	9.930	12,9	890	12,4	9,0
11b	1.801	2,3	175	2,4	9,7
13	5.490	7,1	498	7,0	9,1
Totale	**77.159**	**100,0**	**7.164**	**100,0**	**9,3**

Tabella B.1: Distribuzione degli articoli su rivista nel campione e nella popolazione per ognuno dei GEV bibliometrici

Confrontando le colonne della tabella si nota una buona concordanza tra le percentuali di articoli sottoposti ai GEV nella popolazione e quelli effettivamente referati costituenti il campione empirico: 7.164 articoli. Le cadute riscontrate tra campione teorico e campione empirico sono state oggetto di ulteriore approfondimento facendo ricorso alle tecniche di *Inverse probability weighting*[1] (IPW). Queste tecniche consentono di incorporare le caratteristiche sottostanti il processo di selezione nell'analisi delle stime di effetto pesando le unità (nel nostro caso gli articoli valutati) con la probabilità di selezione, conferendo quindi maggiore stabilità ai risultati ottenuti: in altri termini, offrono uno strumento per valutare gli effetti di selezione non casuale delle unità stesse. Più dettagliatamente, l'ipotesi usata per la costruzione dei pesi è che le unità di analisi siano selezionate in base alle caratteristiche degli articoli (sub-GEV cui l'articolo è stato sottoposto, lingua dell'articolo e classe di valutazione derivante dall'algoritmo bibliometrico[2]). I pesi IPW sono stati costruiti seguendo la procedura cosiddetta *unstabilized* e, successivamente,

[1] Si rimanda, in particolare, a Robins, J.M., Hernán, M.A.; Brumback, B.A. (2000), Marginal Structural Models and Causal Inference in Epidemiology, Epidemiology, 11(5), 550-560; Hernan M.A., Brumback B.A., Robins J.M. (2000), Marginal Structural Models to Estimate the Causal Effect of Zidovudine on the Survival of HIV-Positive Men, Epidemiology, 11(5), 561-570; Hernan M.A., Hernandez-Diaz S., Robins J.M. (2004), A Structural Approach to Selection Bias, Epidemiology, 15(5), 615-625; van der Wal W.M., Geskus R.B. (2011), ipw: An R Package for Inverse Probability Weighting, 43(13), 1-23.

[2] *Proxy* della qualità del prodotto: sede di pubblicazione dell'articolo e impatto in termini di citazioni ricevute.

quella *stabilized*: la prima consiste nel calcolare per ognuno degli articoli del campione teorico la probabilità di essere valutati attraverso un modello logistico che ha come variabile risposta la valutazione (presenza nel campione empirico) e, come covariate, le caratteristiche degli articoli precedentemente menzionate. Il peso viene quindi ottenuto come l'inverso della probabilità predetta dal modello. La seconda procedura (usata per garantire una migliore centratura e simmetria dei pesi intorno al valore uno), consiste nel calcolare l'inverso del rapporto tra le probabilità precedentemente calcolate e quelle predette da un ulteriore modello che usa come unica covariata quella maggiormente predittiva rispetto alla variabile risposta usata (nel nostro caso il sub-GEV). I pesi comunque definiti sono stati usati in un modello di analisi che associa l'esito della *peer*, condotta sugli articoli del campione empirico, con le variabili oggetto di analisi (esito della valutazione bibliometrica ed indicatori bibliometrici); il modello non mostra sostanziali variazioni rispetto ai risultati ottenibili con le stesse variabili, senza effettuare la correzione con i pesi IPW. È quindi ipotizzabile che le discrepanze tra campione teorico e campione empirico non derivino da processi di selezione legati alle caratteristiche degli articoli, ma siano attribuibili a fenomeni casuali verosilmente riconducibili alla mancata collaborazione di alcuni *peer reviewers* o che questo fenomeno di selezione non produca effetti sull'associazione valutazione *peer*/valutazione bibliometrica.

La Tabella B.2 riporta la distribuzione nelle classi di valutazione VQR (Eccellente, Elevato, Discreto, Accettabile, Limitato, Incerto (IR)) della popolazione e del campione stratificato per GEV determinata dalla valutazione bibliometrica degli articoli su rivista. La distribuzione delle valutazioni bibliometriche (Ecc/ El/ D/ A/ L/ IR) è sufficientemente simile nella popolazione e nel campione, per il complesso della VQR e per i singoli GEV ed induce alla conclusione che il campione estratto ben rappresenti la popolazione di riferimento.

GEV 1				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	1.881	40,6	191	43,0
Elevato	994	21,5	99	22,3
Discreto	356	7,7	27	6,1
Accettabile	280	6,0	22	5,0
Limitato	82	1,8	5	1,1
IR	1.038	22,4	100	22,5
GEV 2				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	6.269	61,6	613	60,8
Elevato	1.914	18,8	199	19,7
Discreto	682	6,7	70	6,9
Accettabile	342	3,4	41	4,1
Limitato	48	0,5	3	0,3
IR	927	9,1	82	8,1
GEV 3				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	3.057	46,1	297	45,5
Elevato	1.623	24,5	165	25,3
Discreto	579	8,7	58	8,9
Accettabile	246	3,7	27	4,1
Limitato	29	0,4	2	0,3
IR	1.091	16,5	104	15,9
GEV 4				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	1.110	28,1	98	25,3
Elevato	968	24,5	106	27,3
Discreto	579	14,6	60	15,5
Accettabile	370	9,4	44	11,3
Limitato	99	2,5	12	3,1
IR	827	20,9	68	17,5
GEV 5				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	3.953	37,9	352	37,0

Elevato	2.675	25,7	259	27,2
Discreto	1.256	12,1	108	11,4
Accettabile	673	6,5	64	6,7
Limitato	105	1,0	9	0,9
IR	1.761	16,9	159	16,7
GEV 6				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	6.473	42,0	498	38,5
Elevato	3.395	22,0	296	22,9
Discreto	1.650	10,7	147	11,4
Accettabile	1.150	7,5	101	7,8
Limitato	350	2,3	29	2,2
IR	2.382	15,5	222	17,2
GEV 7				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	1.988	31,3	191	30,3
Elevato	1.701	26,8	164	26,0
Discreto	671	10,6	71	11,3
Accettabile	482	7,6	43	6,8
Limitato	145	2,3	20	3,2
IR	1.367	21,5	141	22,4
GEV 8b				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	969	40,9	99	42,3
Elevato	530	22,4	46	19,7
Discreto	156	6,6	18	7,7
Accettabile	128	5,4	15	6,4
Limitato	28	1,2	2	0,9
IR	559	23,6	54	23,1
GEV 9				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	4.275	43,1	397	44,6
Elevato	2.499	25,2	233	26,2
Discreto	844	8,5	75	8,4
Accettabile	430	4,3	27	3,0
Limitato	76	0,8	7	0,8

IR	1.806	18,2	151	17,0
GEV 11b				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	646	35,9	58	33,1
Elevato	395	21,9	37	21,1
Discreto	191	10,6	13	7,4
Accettabile	118	6,6	19	10,9
Limitato	45	2,5	6	3,4
IR	406	22,5	42	24,0
GEV 13				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	1.980	36,1	153	30,7
Elevato	1.691	30,8	176	35,3
Discreto	886	16,1	80	16,1
Accettabile	514	9,4	45	9,0
Limitato	419	7,6	44	8,8
IR	0	0,0	0	0,0
TOTALE				
Classe	**Popolazione**	**% nella popolazione**	**Campione**	**% nel campione**
Eccellente	32.601	42,3	2.947	41,1
Elevato	18.385	23,8	1.780	24,8
Discreto	7.850	10,2	727	10,1
Accettabile	4.733	6,1	448	6,3
Limitato	1.426	1,8	139	1,9
IR	12.164	15,8	1.123	15,7

Tabella B.2 Distribuzione delle valutazioni bibliometriche nel campione e nella popolazione per ogni GEV e per il complesso della VQR

B.2 Le modalità di confronto

Per ciascun articolo su rivista incluso nel campione casuale sono disponibili le seguenti informazioni:

- valutazione del primo revisore (P1);
- valutazione del secondo revisore (P2);
- valutazione di sintesi dei giudizi del primo e secondo revisore (P);
- valutazione bibliometrica (F).

Le variabili P, P1 e P2 assumono come valore una delle 5 classi di valutazione Ecc, El, D, A, L; la valutazione bibliometrica F ha come possibile risultato anche la classe di valutazione "IR", ossia il suggerimento di procedere con la *informed peer review* nel caso i risultati dei due indicatori bibliometrici (*Impact Factor* e numero di citazioni) siano risultati divergenti. Le cinque classi, secondo il Bando VQR, sono definite con riferimento ai percentili della distribuzione della qualità degli articoli pubblicati nel mondo. In particolare, la qualifica di eccellente corrisponde ad un articolo che si colloca nel primo decile della distribuzione secondo la qualità degli articoli pubblicati nel mondo, quella di elevato nel successivo 20%, quella di discreto nel successivo 20%, accettabile nel successivo 30% e, infine, quella di limitato nel 20% inferiore. Le variabili P1 e P2 sono originariamente misurate su una scala numerica compresa tra 3 e 30, con un punteggio da 1 a 10 assegnato ai 3 criteri fissati nel Bando VQR (originalità, rigore metodologico, impatto attestato o potenziale). Tali punteggi sono successivamente utilizzati per determinare per ciascun prodotto sottoposto a valutazione la classe di valutazione *peer* del prodotto, sulla base dei criteri fissati dal GEV[3]; le variabili P e F sono invece rispettivamente espresse in termini delle 5 (*peer*) o 6 (bibliometria) classi di valutazione sopra elencate. Sulla base del Bando VQR, alle classi Ecc, El, D, A, L corrispondono rispettivamente i punteggi 1; 0,7; 0,4; 0,1; 0.

La classificazione adottata nell'analisi bibliometrica si basa sui criteri descritti nei Rapporti di Area. Nella revisione dei pari, ai revisori esterni è stato richiesto di valutare ciascun prodotto sulla base della loro percezione soggettiva circa la qualità del prodotto rispetto alla distribuzione

[3] L'etichetta "P1" e "P2" assegnata ai revisori è puramente convenzionale e riflette esclusivamente l'ordine di accettazione della proposta di revisione avanzata al potenziale revisore.

mondiale dei prodotti della ricerca nel settore scientifico a cui il prodotto faceva riferimento. La valutazione dei revisori è stata quindi sintetizzata nella valutazione finale sulla base di un algoritmo che confronta la somma dei punteggi attribuiti dai due revisori con quattro soglie[4]. Al fine di confrontare i risultati della valutazione bibliometrica e della revisione tra pari, si procede nel seguito a confrontare gli indicatori F e P. Anche altri confronti possono essere tuttavia d'importanza significativa: in particolare, si utilizzerà anche il confronto tra le valutazioni tra pari P1 e P2 al fine di valutare il grado di corrispondenza dei giudizi tra i due revisori.

Tutte le tabelle mostrate nel seguito derivano dall'analisi del campione di prodotti.

B.3 I risultati

B.3.1 Le distribuzioni della valutazioni F e P

Le distribuzioni delle valutazioni F e P sopra descritte non sono immediatamente confrontabili, in quanto la distribuzione F delle valutazioni bibliometriche comprende una classe IR che non è invece prevista nella valutazione dei pari. È però possibile ipotizzare che una discordanza di almeno due classi tra la valutazione del primo e secondo revisore segnali un'incertezza nella revisione dei pari analoga a quella che emerge dal confronto tra numero di citazioni e fattore d'impatto della sede di pubblicazione nell'analisi bibliometrica. In analogia con la classificazione IR della valutazione bibliometrica, si è creata dunque una classificazione "incerta *peer*" (IP) per la valutazione dei pari, indipendente dalla collocazione nella classe finale, che consente il confronto tra le distribuzioni F e P[5]. La Tabella B.3 mostra la distribuzione in numeri assoluti e percentuali degli indicatori F e P per il totale del campione.

[4] In effetti, ai fini della classificazione VQR, i GEV potevano modificare su base motivata, in alcuni casi anche tramite la richiesta di una terza valutazione *peer*, la classe ottenuta sulla base delle due valutazioni. Per l'esercizio di confronto descritto in questa Appendice, si è escluso l'intervento dei GEV e si è mantenuta la classificazione originaria. Nel caso dell'analisi di confronto effettuata nella VQR 2004-2010 la valutazione di sintesi P era invece riferita alla valutazione finale dei GEV e non alla semplice sintesi aritmetica del punteggio tra i revisori.

[5] I criteri del GEV di Scienze economiche e statistiche non prevedono che l'algoritmo bibliometrico possa giungere a una classe di assegnazione IR. Nel caso del GEV 13 non si è dunque calcolata la classe di assegnazione *peer* IP.

Valutazione bibliometrica (F)	Valutazione peer (P)						
	Eccellente	Elevato	Discreto	Accettabile	Limitato	IP	Totale
Eccellente	755	1.440	322	42	2	386	2.947
Elevato	149	832	480	73	2	244	1.780
Discreto	32	241	254	73	8	119	727
Accettabile	5	113	178	61	10	81	448
Limitato	1	11	43	46	15	23	139
IR	58	436	343	82	9	195	1.123
Totale	**1.000**	**3.073**	**1.620**	**377**	**46**	**1.048**	**7.164**

Tabella B.3a Confronto tra le valutazioni F e P – totale del campione (valori assoluti)

Valutazione bibliometrica (F)	Valutazione peer (P)						
	Eccellente	Elevato	Discreto	Accettabile	Limitato	IP	Totale
Eccellente	25,6	48,9	10,9	1,4	0,1	13,1	100,0
Elevato	8,4	46,7	27,0	4,1	0,1	13,7	100,0
Discreto	4,4	33,1	34,9	10,0	1,1	16,4	100,0
Accettabile	1,1	25,2	39,7	13,6	2,2	18,1	100,0
Limitato	0,7	7,9	30,9	33,1	10,8	16,5	100,0
IR	5,2	38,8	30,5	7,3	0,8	17,4	100,0
Totale	**14,0**	**42,9**	**22,6**	**5,3**	**0,6**	**14,6**	**100,0**

Tabella B.3b Confronto tra le valutazioni F e P – totale del campione (valori % di riga)

Gli elementi sulla diagonale principale della Tabella B.3a corrispondono ai casi in cui la valutazione dei pari e quella bibliometrica coincidono. Gli elementi al di fuori della diagonale principale corrispondono invece ai casi di non coincidenza tra F e P, o perché la valutazione F è migliore della P (elementi al di sopra della diagonale principale) o viceversa (elementi al di sotto della diagonale). La Tabella B.3a mostra che la principale discordanza tra la valutazione bibliometrica e quella dei pari è dovuta al fatto che la valutazione bibliometrica tende a essere più favorevole. In particolare, gli articoli classificati come eccellenti sulla base degli indicatori bibliometrici sono 2.947, quasi il triplo delle valutazioni eccellenti della valutazione tra pari: solo il 25,6% degli articoli classificati come Ecc secondo la bibliometria ottiene Ecc anche secondo la classificazione tra pari, mentre nel 48,9%, 10,9%, 1,4% e 0,1% dei casi gli articoli bibliometricamente eccellenti risultano rispettivamente elevati, discreti, accettabili o limitati nella valutazione dei pari. D'altro lato, il numero di articoli che sono classificati come El e D dalla valutazione tra pari (3.073 e 1.620 articoli rispettivamente) è nettamente più elevato rispetto agli articoli che risultano El e D secondo la valutazione bibliometrica (1.780 e 727 articoli rispettivamente). La numerosità di valutazioni incerte, infine, è quasi uguale nella

revisione tra pari (1.048 articoli) rispetto a quella bibliometrica (1.123 articoli). Le valutazioni bibliometriche incerte hanno nel 38,8% dei casi una valutazione almeno pari a El nell'analisi *peer*, mentre le valutazioni incerte secondo la *peer review* hanno nel 23,3% circa dei casi una valutazione almeno pari a El secondo l'analisi bibliometrica.

Complessivamente, escludendo i casi IR di valutazione incerta (cioè la riga IR e la colonna IP della Tabella B.3a), l'analisi bibliometrica e la revisione tra pari coincidono nel 37% dei casi. Se si sommano alle valutazioni coincidenti quelle che differiscono di una sola classe, si arriva all'87,4% del campione. Gli articoli con valutazioni che differiscono per due classi sono l'11,4% del campione, quelli che differiscono di tre classi sono il 1,2% e, infine, quelli che esibiscono la massima discordanza (ossia, che differiscono per 4 classi) sono lo 0,1%.

La Tabella B.4a mostra la distribuzione degli indicatori P1 e P2. Le valutazioni dei due revisori coincidono nel 37,6% dei casi, sono diverse per una classe di valutazione nel 46,6% dei casi e divergono invece rispettivamente per 2, 3 o 4 classi di valutazione nel 12,7%, nel 2,7% e nello 0,4% dei casi. È da notare anche che le valutazioni su un giudizio di assegnazione alla classe Ecc sono convergenti in 716 casi, pari a circa il 40 % delle valutazioni eccellenti fornite sia dal primo (1.777) che dal secondo (1.723) revisore.

Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	716	743	249	61	8	1.777
Elevato	703	1.269	680	147	37	2.836
Discreto	231	717	570	198	62	1.778
Accettabile	53	174	221	104	37	589
Limitato	20	40	47	42	35	184
Totale	1.723	2.943	1.767	552	179	7.164

Tabella B.4a: Confronto tra le valutazioni P1 e P2 – totale del campione (valori assoluti)

Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	40,3	41,8	14,0	3,4	0,5	100,0
Elevato	24,8	44,7	24,0	5,2	1,3	100,0
Discreto	13,0	40,3	32,1	11,1	3,5	100,0
Accettabile	9,0	29,5	37,5	17,7	6,3	100,0
Limitato	10,9	21,7	25,5	22,8	19,0	100,0
Totale	24,1	41,1	24,7	7,7	2,5	100,0

Tabella B.4b: Confronto tra le valutazioni P1 e P2 – totale del campione (% di riga)

Le Tabelle B.5 e B.6 estendono i risultati delle Tabelle B.3 e B.4 ai singoli GEV. In particolare, dall'analisi dei dati della Tabella B.5 emerge che in tutti i GEV il numero di valutazioni eccellenti è maggiore secondo la valutazione bibliometrica rispetto alla *peer*. Tenuto conto del numero complessivo di articoli valutati in ciascun GEV, la differenza tra il numero di articoli classificati come eccellenti secondo i due metodi di valutazione è particolarmente forte nel GEV6 (498 articoli eccellenti secondo l'algoritmo bibliometrico contro i soli 97 articoli eccellenti secondo l'analisi *peer*), nel GEV9 (397 e 93 articoli eccellenti, rispettivamente secondo l'analisi bibliometrica e quella *peer*) e nel GEV13 (153 e 28 articoli eccellenti, rispettivamente secondo l'analisi bibliometrica e quella *peer)*; le differenze sono invece minori nel GEV1 (191 articoli contro 82) e nel GEV2 (613 articoli rispetto a 318). D'altra parte, il numero di valutazioni elevate è in genere assai maggiore secondo l'analisi *peer* rispetto a quella bibliometrica.

Complessivamente, la tendenza della valutazione bibliometrica a essere più favorevole rispetto a quella *peer* è comune a tutti i GEV (un test statistico che conferma tale ipotesi è presentato successivamente).

GEV 1							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	54	79	23	5	0	30	191
Elevato	12	43	28	2	0	14	99
Discreto	1	5	12	1	0	8	27
Accettabile	0	7	9	2	0	4	22
Limitato	1	0	1	0	0	3	5
IR	14	36	33	4	0	13	100
Totale	82	170	106	14	0	72	444
GEV 2							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	273	262	31	0	0	47	613
Elevato	27	115	36	0	0	21	199
Discreto	3	30	20	2	0	15	70
Accettabile	1	20	11	1	0	8	41
Limitato	0	1	0	0	0	2	3
IR	6	38	26	4	0	8	82
Totale	310	466	124	7	0	101	1.008

GEV 3							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	80	154	26	0	0	37	297
Elevato	15	99	31	1	0	19	165
Discreto	5	21	15	3	0	14	58
Accettabile	1	9	10	3	0	4	27
Limitato	0	1	1	0	0	0	2
IR	9	57	25	2	0	11	104
Totale	110	341	108	9	0	85	653

GEV 4							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	25	46	6	1	0	20	98
Elevato	12	50	24	4	0	16	106
Discreto	4	28	18	1	0	9	60
Accettabile	0	18	13	3	0	10	44
Limitato	0	2	5	3	1	1	12
IR	1	27	22	5	0	13	68
Totale	42	171	88	17	1	69	388

GEV 5							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	86	179	38	1	0	48	352
Elevato	24	129	67	7	0	32	259
Discreto	6	45	34	6	0	17	108
Accettabile	1	12	32	6	0	13	64
Limitato	0	0	3	1	1	4	9
IR	9	70	42	11	0	27	159
Totale	126	435	216	32	1	141	951

GEV 6							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	78	246	71	14	2	87	498
Elevato	12	96	106	16	1	65	296
Discreto	3	40	61	15	1	27	147
Accettabile	1	15	42	18	2	23	101
Limitato	0	4	10	7	3	5	29
IR	3	56	75	23	5	60	222
Totale	97	457	365	93	14	267	1.293

GEV 7							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	39	93	20	6	0	33	191
Elevato	11	82	41	3	0	27	164
Discreto	7	22	30	5	0	7	71
Accettabile	0	9	22	5	0	7	43
Limitato	0	2	7	6	1	4	20
IR	5	53	41	16	3	23	141
Totale	62	261	161	41	4	101	630

GEV 8b							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	16	54	9	2	0	18	99
Elevato	6	26	2	0	0	12	46
Discreto	1	7	5	0	0	5	18
Accettabile	0	5	4	1	0	5	15
Limitato	0	0	2	0	0	0	2
IR	2	27	10	6	0	9	54
Totale	25	119	32	9	0	49	234

GEV 9							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	66	215	51	2	0	63	397
Elevato	16	117	65	4	0	31	233
Discreto	2	26	28	5	1	13	75
Accettabile	1	5	14	2	0	5	27
Limitato	0	1	3	0	0	3	7
IR	8	57	52	9	0	25	151
Totale	93	421	213	22	1	140	890

GEV 11b							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	18	28	8	1	0	3	58
Elevato	6	14	9	1	0	7	37
Discreto	0	5	4	0	0	4	13
Accettabile	0	11	5	1	0	2	19
Limitato	0	0	2	3	0	1	6
IR	1	15	17	2	1	6	42
Totale	25	73	45	8	1	23	175

GEV 13							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	20	84	39	10	0	0	153
Elevato	8	61	71	35	1	0	176
Discreto	0	12	27	35	6	0	80
Accettabile	0	2	16	19	8	0	45
Limitato	0	0	9	26	9	0	44
IR	0	0	0	0	0	0	0
Totale	28	159	162	125	24	0	498

Tabella B.5a: Confronto tra le valutazioni F e P per GEV (valori assoluti)

GEV 1							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	28,3	41,4	12,0	2,6	0,0	15,7	100,0
Elevato	12,1	43,4	28,3	2,0	0,0	14,1	100,0
Discreto	3,7	18,5	44,4	3,7	0,0	29,6	100,0
Accettabile	0,0	31,8	40,9	9,1	0,0	18,2	100,0
Limitato	20,0	0,0	20,0	0,0	0,0	60,0	100,0
IR	14,0	36,0	33,0	4,0	0,0	13,0	100,0
Totale	18,5	38,3	23,9	3,2	0,0	16,2	100,0

GEV 2							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	44,5	42,7	5,1	0,0	0,0	7,7	100,0
Elevato	13,6	57,8	18,1	0,0	0,0	10,6	100,0
Discreto	4,3	42,9	28,6	2,9	0,0	21,4	100,0
Accettabile	2,4	48,8	26,8	2,4	0,0	19,5	100,0
Limitato	0,0	33,3	0,0	0,0	0,0	66,7	100,0
IR	7,3	46,3	31,7	4,9	0,0	9,8	100,0
Totale	30,8	46,2	12,3	0,7	0,0	10,0	100,0

GEV 3							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	26,9	51,9	8,8	0,0	0,0	12,5	100,0
Elevato	9,1	60,0	18,8	0,6	0,0	11,5	100,0
Discreto	8,6	36,2	25,9	5,2	0,0	24,1	100,0
Accettabile	3,7	33,3	37,0	11,1	0,0	14,8	100,0
Limitato	0,0	50,0	50,0	0,0	0,0	0,0	100,0
IR	8,7	54,8	24,0	1,9	0,0	10,6	100,0
Totale	16,8	52,2	16,5	1,4	0,0	13,0	100,0

GEV 4							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	25,5	46,9	6,1	1,0	0,0	20,4	100,0
Elevato	11,3	47,2	22,6	3,8	0,0	15,1	100,0
Discreto	6,7	46,7	30,0	1,7	0,0	15,0	100,0
Accettabile	0,0	40,9	29,5	6,8	0,0	22,7	100,0
Limitato	0,0	16,7	41,7	25,0	8,3	8,3	100,0
IR	1,5	39,7	32,4	7,4	0,0	19,1	100,0
Totale	10,8	44,1	22,7	4,4	0,3	17,8	100,0

GEV 5							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	24,4	50,9	10,8	0,3	0,0	13,6	100,0
Elevato	9,3	49,8	25,9	2,7	0,0	12,4	100,0
Discreto	5,6	41,7	31,5	5,6	0,0	15,7	100,0
Accettabile	1,6	18,8	50,0	9,4	0,0	20,3	100,0
Limitato	0,0	0,0	33,3	11,1	11,1	44,4	100,0
IR	5,7	44,0	26,4	6,9	0,0	17,0	100,0
Totale	13,2	45,7	22,7	3,4	0,1	14,8	100,0

GEV 6							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	15,7	49,4	14,3	2,8	0,4	17,5	100,0
Elevato	4,1	32,4	35,8	5,4	0,3	22,0	100,0
Discreto	2,0	27,2	41,5	10,2	0,7	18,4	100,0
Accettabile	1,0	14,9	41,6	17,8	2,0	22,8	100,0
Limitato	0,0	13,8	34,5	24,1	10,3	17,2	100,0
IR	1,4	25,2	33,8	10,4	2,3	27,0	100,0
Totale	7,5	35,3	28,2	7,2	1,1	20,6	100,0

GEV 7							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	20,4	48,7	10,5	3,1	0,0	17,3	100,0
Elevato	6,7	50,0	25,0	1,8	0,0	16,5	100,0
Discreto	9,9	31,0	42,3	7,0	0,0	9,9	100,0
Accettabile	0,0	20,9	51,2	11,6	0,0	16,3	100,0
Limitato	0,0	10,0	35,0	30,0	5,0	20,0	100,0
IR	3,5	37,6	29,1	11,3	2,1	16,3	100,0
Totale	9,8	41,4	25,6	6,5	0,6	16,0	100,0

GEV 8b							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	16,2	54,5	9,1	2,0	0,0	18,2	100,0
Elevato	13,0	56,5	4,3	0,0	0,0	26,1	100,0
Discreto	5,6	38,9	27,8	0,0	0,0	27,8	100,0
Accettabile	0,0	33,3	26,7	6,7	0,0	33,3	100,0
Limitato	0,0	0,0	100,0	0,0	0,0	0,0	100,0
IR	3,7	50,0	18,5	11,1	0,0	16,7	100,0
Totale	10,7	50,9	13,7	3,8	0,0	20,9	100,0

GEV 9							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	16,6	54,2	12,8	0,5	0,0	15,9	100,0
Elevato	6,9	50,2	27,9	1,7	0,0	13,3	100,0
Discreto	2,7	34,7	37,3	6,7	1,3	17,3	100,0
Accettabile	3,7	18,5	51,9	7,4	0,0	18,5	100,0
Limitato	0,0	14,3	42,9	0,0	0,0	42,9	100,0
IR	5,3	37,7	34,4	6,0	0,0	16,6	100,0
Totale	10,4	47,3	23,9	2,5	0,1	15,7	100,0
GEV 11b							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	31,0	48,3	13,8	1,7	0,0	5,2	100,0
Elevato	16,2	37,8	24,3	2,7	0,0	18,9	100,0
Discreto	0,0	38,5	30,8	0,0	0,0	30,8	100,0
Accettabile	0,0	57,9	26,3	5,3	0,0	10,5	100,0
Limitato	0,0	0,0	33,3	50,0	0,0	16,7	100,0
IR	2,4	35,7	40,5	4,8	2,4	14,3	100,0
Totale	14,3	41,7	25,7	4,6	0,6	13,1	100,0
GEV 13							
Valutazione bibliometrica (F)	**Valutazione peer (P)**						
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**IP**	**Totale**
Eccellente	13,1	54,9	25,5	6,5	0,0	0,0	100,0
Elevato	4,5	34,7	40,3	19,9	0,6	0,0	100,0
Discreto	0,0	15,0	33,8	43,8	7,5	0,0	100,0
Accettabile	0,0	4,4	35,6	42,2	17,8	0,0	100,0
Limitato	0,0	0,0	20,5	59,1	20,5	0,0	100,0
IR	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Totale	5,6	31,9	32,5	25,1	4,8	0,0	100,0

Tabella B.5b: Confronto tra le valutazioni F e P per GEV (valori % di riga)

GEV 1						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	57	39	21	5	2	124
Elevato	47	76	39	5	2	169
Discreto	15	38	41	13	3	110
Accettabile	1	16	15	5	0	37
Limitato	0	2	0	2	0	4
Totale	120	171	116	30	7	444

GEV 2						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	238	161	22	5	0	426
Elevato	153	170	50	8	0	381
Discreto	37	71	44	6	1	159
Accettabile	7	15	12	1	1	36
Limitato	3	2	1	0	0	6
Totale	438	419	129	20	2	1.008

GEV 3						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	84	83	22	3	0	192
Elevato	88	128	63	6	1	286
Discreto	26	68	32	6	2	134
Accettabile	9	14	11	5	0	39
Limitato	0	1	1	0	0	2
Totale	207	294	129	20	3	653

GEV 4						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	27	40	14	4	0	85
Elevato	38	72	47	11	4	172
Discreto	14	31	36	11	6	98
Accettabile	1	9	7	5	1	23
Limitato	1	3	2	3	1	10
Totale	81	155	106	34	12	388

GEV 5						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	84	93	42	9	3	231
Elevato	87	201	93	19	3	403
Discreto	25	106	85	25	5	246
Accettabile	6	17	21	10	2	56
Limitato	4	1	7	3	0	15
Totale	206	418	248	66	13	951

GEV 6						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	65	101	38	9	0	213
Elevato	82	180	143	39	15	459
Discreto	44	146	133	59	20	402
Accettabile	11	49	59	23	12	154
Limitato	4	15	23	12	11	65
Totale	206	491	396	142	58	1.293

GEV 7						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	34	65	15	3	1	118
Elevato	58	110	57	14	2	241
Discreto	30	69	66	20	6	191
Accettabile	3	18	23	15	3	62
Limitato	1	4	4	7	2	18
Totale	126	266	165	59	14	630

GEV 8b						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	20	29	8	4	0	61
Elevato	24	48	21	7	1	101
Discreto	10	22	9	7	3	51
Accettabile	4	8	2	3	0	17
Limitato	1	2	1	0	0	4
Totale	59	109	41	21	4	234

GEV 9						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	70	77	48	8	0	203
Elevato	84	195	100	19	5	403
Discreto	19	104	70	14	4	211
Accettabile	6	21	24	6	3	60
Limitato	4	4	2	2	1	13
Totale	183	401	244	49	13	890
GEV 11b						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	17	21	9	5	0	52
Elevato	14	33	22	4	0	73
Discreto	1	14	18	4	0	37
Accettabile	2	0	3	4	1	10
Limitato	1	0	1	0	1	3
Totale	35	68	53	17	2	175
GEV 13						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	20	34	10	6	2	72
Elevato	28	56	45	15	4	148
Discreto	10	48	36	33	12	139
Accettabile	3	7	44	27	14	95
Limitato	1	6	5	13	19	44
Totale	62	151	140	94	51	498

Tabella B.6a: Confronto tra le valutazioni P1 e P2 per GEV (valori assoluti)

GEV 1						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	46,0	31,5	16,9	4,0	1,6	100,0
Elevato	27,8	45,0	23,1	3,0	1,2	100,0
Discreto	13,6	34,5	37,3	11,8	2,7	100,0
Accettabile	2,7	43,2	40,5	13,5	0,0	100,0
Limitato	0,0	50,0	0,0	50,0	0,0	100,0
Totale	27,0	38,5	26,1	6,8	1,6	100,0

GEV 2						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	55,9	37,8	5,2	1,2	0,0	100,0
Elevato	40,2	44,6	13,1	2,1	0,0	100,0
Discreto	23,3	44,7	27,7	3,8	0,6	100,0
Accettabile	19,4	41,7	33,3	2,8	2,8	100,0
Limitato	50,0	33,3	16,7	0,0	0,0	100,0
Totale	43,5	41,6	12,8	2,0	0,2	100,0

GEV 3						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	43,8	43,2	11,5	1,6	0,0	100,0
Elevato	30,8	44,8	22,0	2,1	0,3	100,0
Discreto	19,4	50,7	23,9	4,5	1,5	100,0
Accettabile	23,1	35,9	28,2	12,8	0,0	100,0
Limitato	0,0	50,0	50,0	0,0	0,0	100,0
Totale	31,7	45,0	19,8	3,1	0,5	100,0

GEV 4						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	31,8	47,1	16,5	4,7	0,0	100,0
Elevato	22,1	41,9	27,3	6,4	2,3	100,0
Discreto	14,3	31,6	36,7	11,2	6,1	100,0
Accettabile	4,3	39,1	30,4	21,7	4,3	100,0
Limitato	10,0	30,0	20,0	30,0	10,0	100,0
Totale	20,9	39,9	27,3	8,8	3,1	100,0

GEV 5						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	36,4	40,3	18,2	3,9	1,3	100,0
Elevato	21,6	49,9	23,1	4,7	0,7	100,0
Discreto	10,2	43,1	34,6	10,2	2,0	100,0
Accettabile	10,7	30,4	37,5	17,9	3,6	100,0
Limitato	26,7	6,7	46,7	20,0	0,0	100,0
Totale	21,7	44,0	26,1	6,9	1,4	100,0

GEV 6						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	30,5	47,4	17,8	4,2	0,0	100,0
Elevato	17,9	39,2	31,2	8,5	3,3	100,0
Discreto	10,9	36,3	33,1	14,7	5,0	100,0
Accettabile	7,1	31,8	38,3	14,9	7,8	100,0
Limitato	6,2	23,1	35,4	18,5	16,9	100,0
Totale	15,9	38,0	30,6	11,0	4,5	100,0

GEV 7						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	28,8	55,1	12,7	2,5	0,8	100,0
Elevato	24,1	45,6	23,7	5,8	0,8	100,0
Discreto	15,7	36,1	34,6	10,5	3,1	100,0
Accettabile	4,8	29,0	37,1	24,2	4,8	100,0
Limitato	5,6	22,2	22,2	38,9	11,1	100,0
Totale	20,0	42,2	26,2	9,4	2,2	100,0

GEV 8b						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	32,8	47,5	13,1	6,6	0,0	100,0
Elevato	23,8	47,5	20,8	6,9	1,0	100,0
Discreto	19,6	43,1	17,6	13,7	5,9	100,0
Accettabile	23,5	47,1	11,8	17,6	0,0	100,0
Limitato	25,0	50,0	25,0	0,0	0,0	100,0
Totale	25,2	46,6	17,5	9,0	1,7	100,0

GEV 9						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	34,5	37,9	23,6	3,9	0,0	100,0
Elevato	20,8	48,4	24,8	4,7	1,2	100,0
Discreto	9,0	49,3	33,2	6,6	1,9	100,0
Accettabile	10,0	35,0	40,0	10,0	5,0	100,0
Limitato	30,8	30,8	15,4	15,4	7,7	100,0
Totale	20,6	45,1	27,4	5,5	1,5	100,0
GEV 11b						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	32,7	40,4	17,3	9,6	0,0	100,0
Elevato	19,2	45,2	30,1	5,5	0,0	100,0
Discreto	2,7	37,8	48,6	10,8	0,0	100,0
Accettabile	20,0	0,0	30,0	40,0	10,0	100,0
Limitato	33,3	0,0	33,3	0,0	33,3	100,0
Totale	20,0	38,9	30,3	9,7	1,1	100,0
GEV 13						
Peer reviewer 1 (P1)	**Peer reviewer 2 (P2)**					
	Eccellente	**Elevato**	**Discreto**	**Accettabile**	**Limitato**	**Totale**
Eccellente	27,8	47,2	13,9	8,3	2,8	100,0
Elevato	18,9	37,8	30,4	10,1	2,7	100,0
Discreto	7,2	34,5	25,9	23,7	8,6	100,0
Accettabile	3,2	7,4	46,3	28,4	14,7	100,0
Limitato	2,3	13,6	11,4	29,5	43,2	100,0
Totale	12,4	30,3	28,1	18,9	10,2	100,0

Tabella B.6b: Confronto tra le valutazioni P1 e P2 per GEV (valori % di riga)

B.3.2 Il confronto tra le distribuzioni di F e P

Il confronto tra la valutazione dei pari e quella bibliometrica si può basare su due criteri fondamentali:

1. il grado di concordanza tra la distribuzione F e la distribuzione P, che analizza la tendenza di F e P ad assegnare lo stesso punteggio a ogni articolo;
2. il grado di differenza esistente tra F e P misurata mediante la differenza media del punteggio assegnato da F e P sulla base dei pesi attribuiti alle classi della VQR.

Ovviamente, una perfetta concordanza implica anche la non esistenza di differenze tra F e P, ma il contrario non è necessariamente vero, e in generale i due criteri misurano due diversi aspetti della differenza esistente tra le due distribuzioni. Si consideri ad esempio una distribuzione con un basso grado di concordanza tra F e P (molti articoli ricevono differenti valutazioni F e P). Anche in tale caso può accadere che, in media, F e P forniscano un punteggio complessivo simile. Questa distribuzione sarebbe caratterizzata da un basso livello di concordanza e da un basso grado di differenza: adottare uno dei due metodi di valutazione (per esempio quella bibliometrica, F) comporterebbe una frequente differenza di valutazione degli articoli sulla base della bibliometria e della valutazione *peer* (ossia, si avrebbero molti articoli con una buona valutazione in base a F, ma una peggiore valutazione in base a P, o viceversa).

Alternativamente, si consideri un caso di elevata (ma non perfetta) concordanza tra F e P. In questo caso, potrebbe ancora succedere che, per esempio, il numero di articoli con classificazione elevata sia sistematicamente maggiore in F che in P. In questo caso si avrebbe un elevato grado di concordanza, ma anche un alto grado di differenza tra le due distribuzioni, dato che il punteggio medio attribuito da F differirebbe dal punteggio medio di P. Adottare uno dei due metodi di valutazione può risultare in una sopravvalutazione (o sottovalutazione) in relazione all'altro criterio: ossia, gli articoli riceverebbero un punteggio notevolmente diverso se valutati con F o con P.

Da un punto di vista statistico, il grado di concordanza tra F e P può essere misurato utilizzando la statistica *K* di Cohen oppure la statistica tau-b di Kendall; differenze tra F e P possono invece essere misurate considerando le differenze tra le medie delle distribuzioni, valutandone la significatività con un test *t* di Student e il test di Wilcoxon (centri delle distribuzioni).

La statistica *K* di Cohen è una misura del grado di concordanza tra giudizi qualitativi espressi sulla base di due diversi metodi o da due diversi revisori; rispetto al semplice calcolo della quota di valutazioni concordanti mostrato in precedenza, *K* tiene conto della possibile concordanza casuale esistente tra i due diversi metodi o revisori. In particolare, *K* è calcolato in modo tale da essere pari a zero quando la concordanza tra le due valutazioni è del tutto casuale, vale a dire nel caso in cui le valutazioni siano indipendenti l'una dall'altra, ed assume invece valore pari a 1 nel caso in cui ci sia perfetta concordanza. Sulla base della stima dell'errore standard ad essa associato, ed assumendo una distribuzione Gaussiana per *K*, ossia approssimando la distribuzione asintotica di *K* con una appropriata distribuzione Gaussiana, è quindi possibile valutare se *K* è statisticamente diverso da zero ad un prescelto livello di confidenza. La statistica

tau-b di Kendall misura invece l'associazione ordinale tra le due quantità, ed è quindi una misura di correlazione tra ranghi.

Quanto al grado di differenza tra le due valutazioni, si calcola in primo luogo la differenza osservata tra le due valutazioni per ciascun articolo e quindi il valor medio delle differenze così calcolate. Si valuta quindi se il valor medio delle differenze tra le due distribuzioni è statisticamente pari a zero, utilizzando una distribuzione *t* di Student; il test *t* si calcola quindi dividendo la media delle differenze per la corrispondente stima della deviazione standard: se il valore ottenuto è superiore al valore soglia della distribuzione *t* di Student corrispondente ad una certo livello prefissato, si conclude che l'evidenza empirica è contro l'ipotesi nulla, ossia che il campione analizzato non è coerente con l'ipotesi che le due valutazioni non provengono da distribuzioni con stessa media. L'equivalente non parametrico del test *t* di Student è il test dei ranghi con segno di Wilcoxon che, a differenza del test *t*, consente di valutare se il valore mediano delle differenze tra le due distribuzioni è diverso da zero senza assumere una distribuzione normale per i punteggi. Il test di Wilcoxon si calcola effettuando la somma dei ranghi corrispondenti alle differenze maggiori di zero. Questo tipo di test è robusto rispetto al test *t* di Student.

B.3.2.1 Il grado di concordanza tra le distribuzioni F e P

La Tabella B.7 riporta i valori della statistica *K* di Cohen, calcolata per l'intero campione e separatamente per ciascun GEV. I risultati sono riferiti a campioni omogenei (*paired sample*), ossia ai prodotti del campione per i quali sono disponibili sia i risultati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione i prodotti per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR; complessivamente, il campione a disposizione si riduce a 6.041 unità/articoli[6]. È possibile calcolare la statistica *K*[7] utilizzando una matrice standard di pesi lineari (1; 0,75; 0,50; 0,25; 0) attribuiti ai casi di concordanza, discordanza di una classe e così via, rispettivamente. In questo caso, nel totale del

[6] Per quanto riguarda l'esito delle valutazioni *peer*, diversamente da quanto descritto all'inizio del paragrafo B.3.1, non si è creata la classificazione "incerta *peer*" (IP), piuttosto l'articolo è stato collocato nella classe finale come previsto dall'algoritmo che confronta la somma dei punteggi attribuiti dai due revisori con quattro soglie.

[7] R Package "vcd", riferimenti: Cohen, J. (1960), A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46. Everitt, B.S. (1968), Moments of statistics kappa and weighted kappa. The British Journal of Mathematical and Statistical Psychology, 21, 97–103. Fleiss, J.L., Cohen, J., and Everitt, B.S. (1969), Large sample standard errors of kappa and weighted kappa. Psychological Bulletin, 72, 332–327.

campione, *K* è uguale a 0,258; la tabella riporta anche la soglia inferiore e quella superiore della stima di *K*, calcolate a partire dalla stima dell'errore standard e per un livello di confidenza del 95%, ipotizzando una distribuzione Gaussiana per *K* . La soglia inferiore della stima di *K* si avvicina in qualche caso allo zero, ma non lo raggiunge mai: è possibile concludere dunque che la concordanza registrata tra la valutazione bibliometrica e quella dei pari non è casuale ma sistematica, ancorché modesta; ossia, le due valutazioni non sembrano tra loro indipendenti. Come accennato sopra, il calcolo di *K* riportato nella prime tre colonne della tabella usa pesi lineari. È possibile argomentare che nel nostro caso i pesi appropriati da utilizzare debbano però essere quelli suggeriti dalle regole della VQR. In particolare, è possibile calcolare la distanza tra le valutazioni utilizzando i punteggi numerici della VQR (1; 0,7; 0,4; 0,1; 0), associati con le valutazioni qualitative (Ecc; El; D; A; L). Le colonne successive della Tabella B.7 riportano i valori della statistica *K* calcolati utilizzando i pesi della VQR. I risultati mostrano che in questo caso la concordanza è maggiore o almeno uguale rispetto alle valutazioni basate su pesi lineari.

GEV	F e P, pesi lineari			F e P, pesi VQR			P1 e P2, pesi lineari			P1 e P2, pesi VQR		
	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K
Totale campione	0,243	0,258	0,274	0,245	0,260	0,276	0,214	0,232	0,250	0,216	0,234	0,252
1	0,150	0,213	0,276	0,151	0,215	0,278	0,126	0,199	0,271	0,129	0,202	0,275
2	0,224	0,264	0,305	0,224	0,265	0,305	0,145	0,194	0,243	0,146	0,195	0,244
3	0,143	0,194	0,246	0,143	0,195	0,247	0,078	0,139	0,200	0,079	0,140	0,201
4	0,165	0,237	0,309	0,166	0,239	0,312	0,110	0,187	0,265	0,111	0,189	0,268
5	0,197	0,240	0,283	0,197	0,241	0,284	0,143	0,192	0,241	0,146	0,195	0,245
6	0,193	0,227	0,260	0,195	0,228	0,262	0,146	0,187	0,227	0,179	0,191	0,232
7	0,210	0,265	0,321	0,213	0,270	0,327	0,132	0,193	0,254	0,134	0,196	0,258
8b	0,088	0,171	0,254	0,088	0,172	0,256	-0,062	0,034	0,129	-0,063	0,035	0,132
9	0,128	0,168	0,208	0,129	0,169	0,210	0,098	0,150	0,202	0,099	0,151	0,203
11b	0,135	0,235	0,334	0,137	0,241	0,345	0,131	0,253	0,376	0,133	0,257	0,382
13	0,257	0,303	0,348	0,262	0,309	0,355	0,257	0,314	0,371	0,254	0,312	0,369

livello di confidenza= 0,95

Tabella B.7: Statistica K di Cohen sul grado di concordanza

La Tabella B.7 riporta anche il coefficiente *K* per il grado di concordanza tra i due revisori (P1 e P2), sia per il totale del campione che per i singoli GEV. Nel complesso del campione, il grado di concordanza tra la valutazione bibliometrica (F) e la revisione *peer* (P) è leggermente

superiore a quello esistente tra i giudizi formulati dai due revisori: in quest'ultimo caso, infatti, la statistica *K* calcolata con i pesi VQR è pari a 0,234 contro 0,260. Analoghi risultati si hanno a livello dei singoli GEV. Anche in questo caso, l'intervallo di variabilità del coefficiente *K* è sempre superiore allo zero con l'unica eccezione del GEV8b. In tutti i 4 casi considerati, il coefficiente *K* assume il valore massimo per il GEV13, anche se lo scostamento non sembra sostanziale.

La Tabella B.8 riporta i valori della statistica tau-b di Kendall, che misura l'associazione ordinale tra due quantità, ed è quindi una misura di correlazione tra ranghi. I valori della tabella mostrano un livello significativo di correlazione esistente tra i ranghi delle due valutazioni F e P, anche in questo caso superiore a quella esistente tra le valutazioni dei due revisori P1 e P2. In particolare, nel caso del GEV13 il coefficiente tau-b, nel caso di F e P, assume il valore di 0,51, nettamente superiore a quello degli altri GEV.

GEV	F e P			P1 e P2		
	soglia inferiore tau-b	tau-b	soglia superiore tau-b	soglia inferiore tau-b	tau-b	soglia superiore tau-b
Totale campione	0,374	0,393	0,412	0,273	0,293	0,313
1	0,231	0,317	0,403	0,172	0,257	0,342
2	0,335	0,383	0,431	0,186	0,242	0,298
3	0,239	0,309	0,379	0,119	0,190	0,261
4	0,255	0,338	0,421	0,155	0,243	0,332
5	0,320	0,374	0,427	0,189	0,247	0,305
6	0,361	0,406	0,450	0,215	0,261	0,307
7	0,281	0,355	0,428	0,199	0,267	0,334
8b	0,133	0,261	0,388	-0,067	0,059	0,185
9	0,252	0,313	0,374	0,182	0,182	0,245
11b	0,237	0,362	0,487	0,140	0,288	0,436
13	0,457	0,509	0,561	0,339	0,401	0,462

Tabella B.8: Statistica tau-b di Kendall sul grado di concordanza

B.3.1.2 Il livello di differenza tra le distribuzioni F e P

La Tabella B.9 riporta il punteggio medio risultante dalle valutazioni F e P. I valori numerici sono ottenuti sommando i pesi assegnati dalla VQR alle 4 classi di merito e dividendo per il numero degli articoli valutati. Si noti ancora una volta come, date le regole della VQR, gli scarti tra F e P non abbiano lo stesso peso: ad esempio, la differenza tra L e A ha un peso pari a 0,1, mentre la differenza tra Ecc ed El ha un peso pari a 0,3. Come nel caso delle analisi contenute nella sezione precedente, i risultati riportati sono riferiti a campioni omogenei (*paired sample*),

ossia ai prodotti del campione per i quali sono disponibili sia i dati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione i prodotti per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR. Come ricordato sopra, gli articoli a disposizione in questo caso sono 6.041.

La quarta colonna mostra che il punteggio medio finale della revisione *peer* (punteggio P) è pari a 0,621: il punteggio è superiore alla media, nell'ordine, in Scienze fisiche, Scienze chimiche, Scienze matematiche e informatiche, Ingegneria civile, Scienze psicologiche, Scienze biologiche e Ingegneria industriale e dell'informazione, ed è invece inferiore nei rimanenti GEV. Le differenze tra i GEV che emergono dall'analisi dei dati della quarta colonna della tabella possono essere attribuite:

- ad una migliore qualità degli articoli sottoposti alla valutazione nei GEV dove il punteggio è superiore alla media complessiva;
- ad una maggiore generosità dei revisori di quei GEV;
- all'intrinseca variabilità statistica nella scelta del campione.

La quinta colonna contiene il punteggio medio ottenuto nella valutazione bibliometrica: tale punteggio è pari a 0,750 per il complesso dei lavori valutabili bibliometricamente, risultando superiore alla media in Ingegneria civile, Scienze matematiche e informatiche, Scienze chimiche, Ingegneria industriale e dell'informazione e Scienze fisiche, in ordine crescente, ed inferiore alla media in Scienze biologiche, Scienze mediche, Scienze psicologiche, Scienze economiche e statistiche e Scienze della terra in ordine decrescente.

L'ordinamento dei GEV in base alla qualità degli articoli presentati secondo la *peer review* vede le Scienze fisiche ricevere le valutazioni migliori, seguite da Scienze chimiche, Scienze matematiche e informatiche, Ingegneria civile, Scienze psicologiche, Ingegneria industriale e dell'informazione e Scienze biologiche; al di sotto della media generale si collocano in ordine decrescente di valutazione le Scienze della terra, Scienze agrarie e veterinarie, Scienze mediche e Scienze economiche e statistiche, che ricevono le valutazioni meno favorevoli.

La sesta colonna della Tabella B.9 presenta la differenza tra valutazione *peer* e bibliometrica, con le colonne 8-11 che riportano il risultato dei test *t* e di Wilcoxon per campioni appaiati. Nel totale del campione, emerge una differenza tra la valutazione bibliometrica e la valutazione *peer*; più precisamente, la valutazione media ottenuta con l'analisi bibliometrica è superiore rispetto a quella ottenuta con la valutazione *peer*. Il risultato è confermato anche dai dati riferiti ai singoli

GEV: fanno eccezione i GEV di Scienze della terra e Scienze Psicologiche, nei quali la differenza tra la valutazione dei pari e quella bibliometrica non è statisticamente significativa agli usuali livelli di confidenza.

GEV	Punteggio P1	Punteggio P2	Punteggio P	Punteggio F	Diff F-P	# Osservazioni	Test t	*p*-value di t	Test di Wilcoxon	*p*-value di Wilcoxon
1	0,663	0,662	0,648	0,794	0,147	344	7,424	0,000	19211,0	0,000
2	0,768	0,789	0,751	0,847	0,097	926	9,061	0,000	98818,0	0,000
3	0,698	0,724	0,683	0,799	0,115	549	8,182	0,000	46441,5	0,000
4	0,659	0,619	0,612	0,627	0,015	320	0,680	0,497	11374,0	0,330
5	0,658	0,651	0,632	0,736	0,104	792	7,937	0,000	98681,5	0,000
6	0,570	0,572	0,546	0,723	0,177	1.071	14,333	0,000	246306,5	0,000
7	0,611	0,626	0,596	0,692	0,097	489	5,419	0,000	33218,5	0,000
8b	0,666	0,692	0,643	0,777	0,134	180	4,935	0,000	6025,0	0,000
9	0,671	0,645	0,630	0,802	0,172	739	14,224	0,000	109976,5	0,000
11b	0,701	0,638	0,641	0,684	0,043	133	1,169	0,244	2358,0	0,205
13	0,483	0,468	0,435	0,628	0,193	498	9,849	0,000	56761,0	0,000
Totale	0,648	0,647	0,621	0,750	0,129	6.041	25,924	0,000	6072973,0	0,000

Tabella B.9: Test t e Wilcoxon sulla differenza tra i punteggi bibliometrici e peer review

Infine, nella Tabella B.10 sono riportati i valori del coefficiente di correlazione tra diversi indicatori utilizzabili per la classificazione finale degli articoli. La descrizione degli acronimi è mostrata nella seconda tabella. Gli indicatori considerati sono: il numero di autori di ciascuna pubblicazione (acronimo: N_A); l'indicatore di impatto della rivista sede di pubblicazione dell'articolo (J_I); il corrispondente percentile dell'indicatore di impatto della rivista usato dall'algoritmo bibliometrico VQR per l'attribuzione della classe di merito (J_p); il numero di citazioni dell'articolo (C_N); il corrispondente percentile di citazioni dell'articolo usato dall'algoritmo bibliometrico VQR per l'attribuzione della classe di merito (Ct_); il punteggio somma dei giudizi espressi dal primo *peer reviewer* rispetto ai tre criteri di valutazione fissati nel bando (range 3:30) (P1_); il punteggio somma dei giudizi espressi dal secondo *peer reviewer* rispetto ai tre criteri di valutazione fissati nel bando (range 3:30) (P2_); la classe dell'algoritmo VQR espressa in punteggio finale VQR (1; 0,7; 0,4; 0,1; 0) (A_V); la valutazione sintetica dei due *peer reviewers* ed espressa in punteggio finale VQR[8] (P_V); il punteggio complessivo dei tre criteri di valutazione fissati nel bando espresso come somma dei punteggi di entrambi i *peer*

[8] Come spiegato nel paragrafo B.2.

reviewers (range 6-60). La Figura 1 presenta una visualizzazione grafica delle associazioni tra gli indicatorimenzionati. Se si utilizza come *benchmark* la valutazione sintetica dei due *peer reviewers* espressa in punteggio finale VQR (P_V), si può notare come essa sia maggiormente correlata con la classe dell'algoritmo VQR (espressa nel corrispondente punteggio finale VQR), piuttosto che con i due singoli indicatori bibliometrici utilizzati (Cit_perc e Jou_perc), a riprova del fatto che entrambi forniscono un contributo significativo alla valutazione bibliometrica e, insieme, la rendono più vicina a quella *peer*.

Variabili	Acronimi	N Autori	Jou Ind	Jou perc	Cit Num	Cit perc	P1 VQR score	P2 VQR score	Classe algoritmo VQR	P VQR	P score
N Autori	**N_A**	1,000	0,010	0,090	0,388	0,145	0,144	0,157	0,133	0,181	0,188
Jou Ind	**J_I**	0,010	1,000	0,299	0,116	0,257	0,217	0,220	0,275	0,258	0,267
Jou perc	**J_p**	0,090	0,299	1,000	0,079	0,843	0,209	0,209	0,388	0,229	0,265
Cit Num	**C_N**	0,388	0,116	0,079	1,000	0,137	0,121	0,123	0,132	0,152	0,161
Cit perc	**Ct_**	0,145	0,257	0,843	0,137	1,000	0,234	0,235	0,587	0,263	0,295
P1 VQR score	**P1_**	0,144	0,217	0,209	0,121	0,234	1,000	0,307	0,371	0,734	0,773
P2 VQR score	**P2_**	0,157	0,220	0,209	0,123	0,235	0,307	1,000	0,370	0,713	0,758
Classe algoritmo VQR	**A_V**	0,133	0,275	0,388	0,132	0,587	0,371	0,370	1,000	0,424	0,464
P VQR	**P_V**	0,181	0,258	0,229	0,152	0,263	0,734	0,713	0,424	1,000	0,917
P score	**P_s**	0,188	0,267	0,265	0,161	0,295	0,773	0,758	0,464	0,917	1,000

Tabella B.10. Matrice di correlazione tra diversi indicatori di valutazione

Nome variabile	Acronimo / Nome nodo del grafo	Descrizione variabili
N Autori	**N_A**	Numero di autori
Jou Ind	**J_I**	Indicatore della Rivista
Jou perc	**J_p**	Percentile dell'indicatore della Rivista
Cit Num	**C_N**	Numero di citazioni
Cit perc	**Ct_**	Percentile di citazioni
P1 VQR score	**P1_**	Score P1 punteggio VQR
P2 VQR score	**P2_**	Score P2 punteggio VQR
Classe algoritmo VQR	**A_V**	Algoritmo espresso in punteggio VQR
P VQR	**P_V**	P espresso come punteggio VQR
P score	**P_s**	P espresso come somma punteggi criteri

Legenda Tabella B.10 e Figura 1.

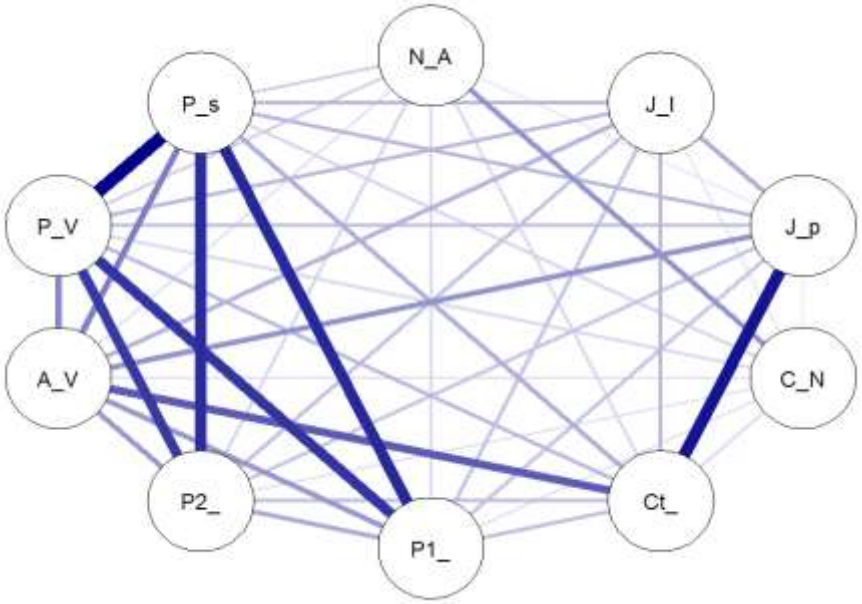

Figura 1. Grafo pesato della Tabella B.10

B.4 Conclusioni

L'analisi effettuata suggerisce la presenza di una correlazione non nulla tra la valutazione *peer* e la valutazione bibliometrica. Il coefficiente *K* di Cohen indica un grado di concordanza modesto, ancorché significativo. Il coefficiente tau-b di Kendall, che misura la correlazione tra ranghi, assume valori più elevati del coefficiente *K* per tutti i GEV e, in particolare, è maggiore di 0,5 per il GEV 13[9].

[9] Una possibile spiegazione di questo risultato può risiedere nella diversità dell'algoritmo usato dal GEV 13 per la valutazione bibliometrica. Come noto dalla pubblicazione del documento *Criteri per la valutazione dei prodotti di ricerca*, il GEV 13 ha indicato una classificazione delle riviste separata per ogni sotto-area usando un algoritmo che ha combinato esclusivamente gli indicatori bibliometrici di impatto delle riviste forniti da ISI WoS, Scopus e Google Scholar. Le citazioni sono state quindi ignorate nella classificazione diretta delle riviste, ma potevano comunque garantire una classe immediatamente superiore agli articoli pubblicati su riviste indicizzate che nel periodo 2011-2014 avessero presentato un numero di citazioni annuali medie superiore all'indicatore di impatto (misurato con l'indicatore IF5Y di ISI WoS o l'indicatore IPP di Scopus) di quella data rivista nel 2014.
Per maggiori dettagli sul processo di valutazione bibliometrica adottato dal GEV 13 si rimanda al documento criteri del GEV ed ai comunicati pubblicati sul sito ANVUR al seguente url:

È di particolare importanza il risultato che il grado di concordanza (di classe e tra ranghi) tra la valutazione bibliometrica e quella *peer* sia sempre superiore a quello esistente tra le due valutazioni *peer* individuali.

Emerge però in tutti i GEV l'evidenza di differenze di segno positivo tra i punteggi medi corrispondenti alle valutazioni *peer* e bibliometriche (ossia, la valutazione bibliometrica è significativamente più favorevole in media rispetto a quella *peer*). In effetti, la percentuale di prodotti della ricerca classificati come eccellenti (Ecc) con l'algoritmo di valutazione bibliometrica è sempre superiore a quello dei prodotti eccellenti secondo la valutazione tra pari.

La differenza tra peer review e valutazione bibliometrica in termini di valutazione media e, soprattutto, nella maggiore percentuale di valutazioni eccellenti per la valutazione bibliometrica, non deve stupire, anche se si è cercato il più possibile di evitare questo esito chiedendo ai GEV la massima coerenza con i criteri bibliometrici. La *peer review* è infatti una metodologia di valutazione soggettiva, ed è prassi comune che i revisori attribuiscano in prevalenza valutazioni intermedie (elevato, discreto e accettabile) piuttosto che estreme.

Nella bibliometria, invece, i criteri seguiti dai GEV si sono strettamente basati sulla definizione del Bando (10%, 20%, 20%, 30% e 20%) ed inoltre la pubblicazione dei criteri precedente al conferimento dei prodotti ha consentito alle strutture di scegliere i prodotti da valutare in base a criteri precisi, che nella maggior parte dei casi ne garantivano il risultato (ad esempio la classificazione in Ecc).

---

http://www.anvur.it/index.php?option=com_content&view=article&id=856:area-13-scienze-economiche-e-statistiche-2011-2014&catid=2:non-categorizzato&Itemid=616&lang=it.