Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Appendice A

Comparison between peer review and bibliometric evaluation

1. The random sample

A sample of 590 journal articles was randomly drawn from the population of 5,681 articles published in the journals classified by GEV13. The sample was stratified by research area and includes 10% of the journal articles for areas E (Economics), M (Management), S (Statistics) and 25% for area H (History). The sample was drawn in early September 2012, before starting the peer review process, by selecting one every ten articles (one every four in area H) by a random number generator.

Table 1. Distribution of journal articles in the population and in the sample

	Population	*Sample*	*%*
Economics	2361	235	10
History	147	37	25
Management	1750	175	10
Statistics	1423	143	10
Total	5681	590	

Table 2. Distribution of bibliometric ranking in the population and in the sample

	N Population	*% Population*	*N Sample*	*% Sample*
Economics				
A	923	39.09	95	40.43
B	337	14.27	29	12.34
C	434	18.38	49	20.85
D	667	28.25	62	26.38
History				
A	35	23.81	9	24.32
B	43	29.25	12	32.43
C	25	17.01	7	18.92
D	44	29.93	9	24.32
Management				
A	458	26.28	44	25.14
B	238	13.65	22	12.57
C	231	13.25	31	17.71
D	816	46.82	78	44.57
Statistics				
A	507	35.63	51	34.97
B	382	26.84	38	27.27
C	166	11.67	16	11.19
D	368	25.86	37	26.57

Table 1 reports the distribution of the journal articles in the 4 areas, in the population and in the sample. Table 2 shows that the population and the sample distributions of the bibliometric ranking (A / B / C / D) are very close for each research area. In other words, our sample is representative of the population of journal articles both overall and within each area.

For each journal article included in the random sample the following variables are available:

- report by the first referee (P1);
- report by the second referee (P2);
- evaluation by the Consensus Group (P);
- bibliometric indicator (F).

Each of these four qualitative variables is mapped into one of four classes (A, B, C, D), corresponding respectively to the top 20% of the quality distribution of published papers, the next 20%, the next 10%, and the bottom 50%. More precisely, variables P1 and P2 are originally measured on a numerical scale between 3 and 27, with scores from 1 to 9 assigned to 3 different criteria, and are then converted into one of the above 4 levels using the VQR grid;[1] the other two are directly expressed in the 4-level format. According to the VQR rules, the 4 classes correspond to scores 1, 0.8, 0.5 and 0, respectively.

The classification adopted in the bibliometric analysis is derived from the GEV13 journal ranking, which is based on the 5-year Impact Factor (IF5) and the Article Influence Score (AIS) indicators (imputed for journals with missing values for these variables), adjusted for highly cited papers (more than 5 citations per year). The appropriate quantiles of the distribution are then mapped into the 4 classes A-D. In the peer review, two external referees (P1 and P2) were asked to rate each product according to the world-wide quality distribution as they subjectively perceived it. The opinion of the external referees was then summarized by the internal Consensus Group: in cases of disagreement between P1 and P2, the P index is not simply the average of P1 and P2, but also reflects the opinion of two (and occasionally three) members of GEV13 (as described in detail in the document devoted to the peer review process).

To compare peer review and bibliometric analysis, one should compare the F and P indicators. However, other comparisons are also informative. In particular, the comparison between P1 and P2 allows us to study how much agreement there is between the two external referees.

[1] Labeling the two referees as “P1” or “P2” is purely a convention, reflecting only the order in which the referees accepted to review the paper.

2. The F and P distributions

Table 3 tabulates the distribution of the F and P indicators, Table 4 tabulates the distribution of P1 and P2, while Table 5 tabulates the distribution of F and P separately by research areas.

The elements on the main diagonal of Table 3 correspond to cases for which peer review and bibliometric evaluations coincide. The off-diagonal elements correspond to cases of disagreement between P and F, either because F provides a higher evaluation (elements above the main diagonal) or because P provides a higher evaluation (elements below the main diagonal).

Table 3 shows that the main source of disagreement between F and P is the fact that peer review classifies less papers in "A" (only 116) compared to bibliometric analysis (198 papers). Of the 198 papers classified as "A" by the bibliometric analysis, peer review classifies as "A" only 49% of them. The table also shows that a larger number of papers are classified as "B" (174) by the peer review compared to bibliometric analysis (102 papers). On the other hand, the assignment of papers to the "C" and "D" classes is similar for peer review and bibliometric analysis. In summary, Table 3 shows that bibliometric analysis tends to be more generous than peer review for "A" papers.

Table 3. Comparison between F and P – Total sample

	Peer (P)				
Bibliometric (F)	A	B	C	D	Total
A	98	72	19	9	198
	49.49	36.36	9.60	4.55	100.00
B	11	56	26	9	102
	10.78	54.90	25.49	8.82	100.00
C	4	25	39	35	103
	3.88	24.27	37.86	33.98	100.00
D	3	21	45	118	187
	1.60	11.23	24.06	63.10	100.00
Total	116	174	129	171	590
	19.66	29.49	21.86	28.98	100.00

In the total sample, bibliometric analysis (F) and peer review (P) give the same classifications in 53% of the cases (311 cases are on the main diagonal of Table 6) , and in 89% of the cases differ by at most one class. Extreme disagreement (difference of 3 classes) occurs in only 2% of the cases, and a milder form of disagreement (difference of 2 classes) in only 9% of the cases.

Table 4. Comparison between P1 and P2

	Peer #2				
Peer #1	**A**	**B**	**C**	**D**	**Total**
A	53	43	7	11	114
	46.49	37.72	6.14	9.65	100.00
B	36	73	29	29	167
	21.56	43.71	17.37	17.37	100.00
C	8	34	21	29	92
	8.70	36.96	22.83	31.52	100.00
D	4	46	50	117	217
	1.84	21.20	23.04	53.92	100.00
Total	101	196	107	186	590
	17.12	33.22	18.14	31.53	100.00

Table 4 cross-tabulates the opinion of the two external referees. In 45% of the cases they agree on the same evaluation, and in 82% of the cases their evaluation differs by at most one class. Note also that referees agree on an "A" evaluation in about half of the cases.

Table 5 cross-tabulates F and P by research area. Disagreement by more than one class occurs in 19% of the cases for History, but only in 10% of the cases for the other three areas. The lower frequency of "A" and the higher frequency of "B" in the peer review, relative to the bibliometric analysis, occurs for all areas except History, where 10 papers are classified as "A" by the peer review and 9 by the bibliometric analysis. In this case, however, the sample is relatively small (37 observations) and cell-by-cell comparison might not be reliable.

Table 5. Comparison between B and P, by area

Economics		*Peer*			
Bibliometric (F)	**A**	**B**	**C**	**D**	Total
A	48	39	8	0	95
	50.53	41.05	8.42	0.00	100.00
B	2	12	10	5	29
	6.90	41.38	34.48	17.24	100.00
C	1	8	22	18	49
	2.04	16.33	44.90	36.73	100.00
D	1	9	11	41	62
	1.61	14.52	17.74	66.13	100.00
Total	52	68	51	64	235
	22.13	28.94	21.70	27.23	100.00

History		*Peer*			
Bibliometric (F)	**A**	**B**	**C**	**D**	**Total**
A	5	3	0	1	9
	55.56	33.33	0.00	11.11	100.00
B	4	7	0	1	12
	33.33	58.33	0.00	8.33	100.00
C	1	3	2	1	7
	14.29	42.86	28.57	14.29	100.00
D	0	4	3	2	9
	0.00	44.44	33.33	22.22	100.00
Total	10	17	5	5	37
	27.03	45.95	13.51	13.51	100.00

Management		*Peer*			
Bibliometric (F)	**A**	**B**	**C**	**D**	**Total**
A	17	11	9	7	44
	38.64	25.00	20.45	15.91	100.00
B	0	13	7	2	22
	0.00	59.09	31.82	9.09	100.00
C	1	9	7	14	31
	3.23	29.03	22.58	45.16	100.00
D	0	4	17	57	78
	0.00	5.13	21.79	73.08	100.00
Total	18	37	40	80	175
	10.29	21.14	22.86	45.71	100.00

Statistics		*Peer*			
Bibliometric (F)	**A**	**B**	**C**	**D**	**Total**
A	28	19	2	1	50
	56.00	38.00	4.00	2.00	100.00
B	5	24	9	1	39
	12.82	61.54	23.08	2.56	100.00
C	1	5	8	2	16
	6.25	31.25	50.00	12.50	100.00
D	2	4	14	18	38
	5.26	10.53	36.84	47.37	100.00
Total	36	52	33	22	143
	25.17	36.36	23.08	15.38	100.00

3. Comparison between F and P

When comparing peer review and bibliometric analysis two criteria could be considered:

1. Amount of a**greement** between F and P, that is, if F and P tend to agree on the same score.
2. **Systematic difference** between F and P, that is, the average of the score difference between F and P.

Of course, perfect agreement would imply no difference, but the opposite is not true and in general the two criteria highlight somewhat different aspects. Consider for instance a distribution with a large amount of disagreement between F and P (many papers receive different evaluations according to the F and P criteria). It could still be that, on average, F and P provide a similar evaluation. This distribution is characterized by **low agreement and low differences**. Adopting one of the two evaluations (for instance the F evaluation) would result in frequent misclassification of papers according to the other criterion (e.g., many papers with good F but poor P evaluations, and vice versa).

Alternatively, consider a case of close (but not perfect) agreement between F and P. It could still be that, for instance, F assigns a higher class more often than P. This distribution is characterized by **high agreement but also large differences**, as the average F score differs from the average P score in a systematic way. Adopting one of the two evaluations would result in over-evaluation (or under-evaluation) if measured with reference to the other criterion; that is, on average papers receive a higher (or a lower) score using the F or P evaluations

From the statistical point of view, the amount of agreement between F and P may be measured using Cohen's kappa, while systematic differences between sample means can be detected using a standard *t*-test for paired samples.

4. The degree of agreement

Table 6 reports the kappa statistic for the entire sample and separately for each research area. The kappa statistic is scaled to be 0 when the amount of agreement is what one would expect to observe by pure chance and 1 when there is perfect agreement. The statistic is computed using standard linear weights (1, 0.67, 0.33, 0), to take into account that cases of mild disagreements (say, disagreement between "A" and "B") should weight less than cases of stronger disagreements (say, disagreements between "A" and "C", and between "A" and "D").

In the total sample, kappa is equal to 0.54 and is statistically different from zero at the 1% level. The agreement for Economics, Management and Statistics is close to the overall sample agreement, while History features a lower value of kappa (0.32). For each research area, kappa is statistically different from zero at the 1% level.

As mentioned, the computation of kappa in the first row of Table 6 uses linear weights. It can be argued that, in the present context, the appropriate weights are those of the VQR. These weights

compute the distance between the evaluations using the numerical scores (1, 0.8, 0.5, 0) associated with the qualitative evaluations (A, B, C, D). The second row of Table 6 reports the "VQR weighted" kappa. The statistic is quite similar to the linearly weighted kappa, indicating good agreement in the total sample (0.54) and in three research areas (Economics, Management, and Statistics), and a lower value in History (0.29).

Table 6. Kappa statistic for the amount of agreement

	Total sample (1)	Economics (2)	History (3)	Management (4)	Statistics (5)
F and P, linear weighted kappa	0.54 (18.11)**	0.56 (11.94)**	0.32 (2.95)**	0.49 (8.91)**	0.55 (9.41)**
F and P, VQR weighted kappa	0.54 (17.29)**	0.56 (11.53)**	0.29 (2.56)**	0.50 (8.37)**	0.55 (9.18)**
P1 and P2, equal weights	0.40 (12.93)**	0.44 (9.06)**	0.18 (1.49)	0.33 (5.90)**	0.33 (5.47)**
P1 and P2, VQR weights	0.39 (12.06)**	0.42 (8.28)**	0.15) (1.29)	0.33 (5.55)**	0.32 (5.17)**

Note. The table reports the kappa statistic and the associated z-value in parenthesis. One star indicates significance at the 5% level; two stars indicate significance at the 1% level.

The degree of agreement between bibliometric ranking (F) and peer review (P) is higher than that between the two external referees (P1 and P2). To see this, Table 6 reports the kappa statistics for the degree of agreement between the two referees (P1 and P2) in the total sample and separately by research area. In the total sample, the linearly weighted kappa is equal to 0.40 (0.39 using VQR weights), and is lower than the corresponding kappa for the comparison of F and P (0.54 for both the linear and the VQR weights). For each research area, the pattern is similar to that observed when comparing F and P. For Economics, Management and Statistics there is more agreement between referees than for History (for this area, kappa is not statistically different from zero). Furthermore, for each research area there is more agreement between F and P than between P1 and P2.

5. Systematic differences

Table 7 reports the average score resulting from the F and P evaluations. Numerical scores are obtained converting the qualitative F and P evaluations (A, B, C or D) using the weights assigned by the VQR to the four merit classes (1, 0.8, 0.5, 0). Note once more that, given the rules of the VQR, deviations between F and P do not carry the same weight: for instance, a difference between "D" and "C" has a weight of 0.5, while a difference between "A" and "B" has a weight of only 0.2.

Table 7 reports the average numerical score of the two referees (columns labeled "Score P1" and "Score P2"). Column 3 reports that the average score of the peer review ("Score P") is equal to 0.542. The score is lower for Management (0.386) and higher for History (0.706) and Statistics (0.658). The difference across areas in column 3 might arise for several reasons, including:

- the pool of papers in History and Statistics is of higher quality;
- referees in History and Statistics tend to be more generous than in other areas;
- sampling variability.

Column 4, labeled "F", reports the average score of the bibliometric evaluation (0.561). As the P score, the F score tends to be lower in Management (0.444) and slightly higher in Statistics (0.624).

The most interesting statistic can be seen in column 5, which reports the difference between F and P scores, and the associated paired *t*-statistic in column 7. In the overall sample the difference is positive (0.019) and not statistically different from zero at conventional levels (the *p*-value is 0.157). However, there are differences across areas. For Economics and Management, the difference is positive (0.046 and 0.057, respectively) and statistically different from zero at the 5% level (but not at the 1% level). For Statistics and History, the difference is negative (-0.108 and -0.034, respectively) but not statistically different from zero.

Table 7. Test for the difference between average B and P scores

	Score P1	Score P2	Score P	Score F	Difference between F and P	Sample size	t-test for difference between F and B	p-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
E	0.503	0.521	0.561	0.607	0.046	235	2.286	0.023
H	0.649	0.700	0.705	0.597	-0.108	37	-1.672	0.103
M	0.335	0.421	0.386	0.441	0.054	175	1.999	0.047
S	0.649	0.625	0.658	0.624	-0.034	143	-1.417	0.159
Total	0.498	0.528	0.542	0.561	0.019	590	1.417	0.157

Note. The F and P scores are obtained using the values: A=1; B=0.8; C=0.5; D=0. The t-test is computed for paired samples.

6. Assessment

In the total sample there is more than adequate agreement between F and P. Furthermore, there is no evidence of systematic differences between the average scores provided by the F and P rankings. Although in the aggregate there are no systematic differences between F and P, there is a lower number of papers classified by referees as "A" relative to the bibliometric analysis. However, most of the papers "downgraded" by the peer review are still classified as "B", and deviations from the two upper classes do not carry a large weight in the VQR.

For three of the four research areas, there is more than adequate agreement, while the agreement is somewhat lower for History. Systematic differences between the average scores for the four research areas are generally small and not always of the same sign: for Economics and Management the difference is positive and statistically significant at the 5% level, while for Statistics and History the difference is negative and not statistically different from zero.

There are at least three possible reasons that can account for heterogeneity in the four research areas:

- *Differences among referees*. Referees are less generous in Economics or Management than in Statistics and especially History.
- *Reliability of the journal ranking*. The ranking of journals is more generous (for instance a larger number of journals in the highest classes) in Economics and Management relative to other areas.
- *Power of the test*. When the sample size is not large, as it is true in some research areas, confidence intervals tend to be relatively large.

Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Appendice B
Criteri per la Valutazione dei Prodotti di Ricerca

Gruppo di Esperti della Valutazione dell'Area Scienze Economiche e Statistiche (GEV 13)

1. Introduzione

Questo documento descrive l'organizzazione del Gruppo di Esperti della Valutazione di Scienze Economiche e Statistiche (d'ora in poi, GEV) ed i criteri che il Gruppo utilizzerà per valutare i prodotti di ricerca. Dopo l'introduzione, il documento si divide in 4 parti. La sezione 2 elenca le aree di ricerca del GEV. La sezione 3 riassume le regole interne. La sezione 4 descrive i due strumenti di valutazione (bibliometrico e *peer review*) e spiega come il GEV selezionerà i prodotti di ricerca che saranno valutati in *peer review*. In particolare la sezione 4.1 descrive il processo di *peer review*, le linee guida per la scelta dei revisori esterni e le procedure del GEV per risolvere i conflitti di valutazione. La sezione 4.2 descrive come il GEV affronterà l'analisi bibliometrica, e come selezionerà l'elenco delle riviste e gli indicatori bibliometrici per gli articoli su rivista. Infine, la sezione 5 descrive come il GEV programma di risolvere i conflitti di interesse tra i componenti del GEV e gli autori dei prodotti di ricerca.

2. Le aree di ricerca

	Area 13: Scienze Economiche e Statistiche Settori scientifico-disciplinari (SSD) di riferimento
SECS/P01	Economia politica
SECS/P02	Politica economica
SECS/P03	Scienza delle finanze
SECS/P04	Storia del pensiero economico
SECS/P05	Econometria
SECS/P06	Economia applicata
SECS/P07	Economia aziendale
SECS/P08	Economia e gestione delle imprese
SECS/P09	Finanza aziendale
SECS/P10	Organizzazione aziendale
SECS/P11	Economia degli intermediari finanziari
SECS/P12	Storia economica
SECS/P13	Scienze merceologiche
SECS/S01	Statistica
SECS/S02	Statistica per la ricerca sperimentale e tecnologica
SECS/S03	Statistica economica
SECS/S04	Demografia
SECS/S05	Statistica sociale
SECS/S06	Metodi matematici dell'economia e delle scienze attuariali e finanziarie

Tabella 1. I settori scientifico- disciplinari (SSD) di riferimento dell'Area 13 Scienze Economiche e Statistiche

3. Organizzazione del GEV

Coordinatore del GEV: Tullio Jappelli

3.1. Sub-GEV

Nome del SUB-GEV e aree di ricerca (SSD)	Coordinatore	Componenti
Economia SECS/P01 SECS/P02 SECS/P03 SECS/P04 SECS/P06 SECS/P12	Graziella Bertocchi	Alberto Bisin Laura Bottazzi Giorgio Brunello Fabio Canova Daniele Checchi Valentino Dardanoni Daniela Del Boca Giovanni Dosi Leonardo Felli Fabiano Schivardi Neri Salvadori Vera Zamagni
Economia Aziendale e Finanza SECS/P07 SECS/P08 SECS/P09 SECS/P10 SECS/P11 SECS/P13	Alfonso Gambardella	Massimo Bergami Angelo Cichelli Francesca Cornelli Andrew Ellul Alex Frino Gianluigi Guido Giovanni B. Dagnino James Guthrie Maurizio Murgia Riccardo Mussari Paolo Quattrone Daniele Terlizzese Massimo Warglien
Statistica e Matematica Applicata SECS/P05 SECS/S01 SECS/S02 SECS/S03 SECS/S04 SECS/S05 SECS/S06	Franco Peracchi	Francesco Bartolucci Marco Frittelli Marc Hallin Massimo Marinacci Elvezio Ronchetti Barbara Rossi Guglielmo Weber

Tabella 2. Sub-GEV, corrispondenti settori scientifico- disciplinari (SSD), coordinatori e componenti

Ciascun componente assegnato ad un sub-GEV può essere successivamente assegnato ad altri sub-GEV sulla base della sua area di ricerca e della necessità degli altri sub-GEV. Il coordinatore del GEV (Tullio Jappelli) non viene assegnato ad alcun sub-GEV, e collaborerà con gli altri GEV nella valutazione dei prodotti di ricerca multidisciplinari.

3.2. Distribuzione dei prodotti di ricerca

La distribuzione dei prodotti ai sub-GEV avverrà sulla base del SSD indicato dal soggetto valutato così come trasmesso dalle strutture. Il significato del SSD associato al prodotto, che può anche essere diverso dal SSD di appartenenza del soggetto valutato, si riferisce al GEV che con maggior competenza, secondo il soggetto valutato, può valutare il prodotto in questione.

Nel caso in cui un prodotto sia assegnato a più GEV per il suo carattere interdisciplinare saranno adottati identici criteri di valutazione concordati tra i vari GEV. A tale scopo, i Presidenti dei GEV interessati istituiscono specifici Gruppi di Consenso Inter-Area.

3.3. Regole interne

- La convocazione del GEV avviene almeno 15 giorni prima della riunione. La riunione è convocata dal Presidente, che fissa anche l'ordine del giorno.

- Le decisioni all'interno del GEV vengono prese a maggioranza semplice. Per partecipare alla votazione non è necessario essere fisicamente presenti alle riunioni, purché presenti in modalità telematica.

- Alle riunioni del GEV partecipa, con funzioni di segretario senza diritto di voto, anche l'assistente al GEV, dott.ssa Carmela Anna Nappi, attribuito da ANVUR al GEV. Al termine di ciascuna riunione viene redatto un resoconto della seduta in italiano, e un verbale sintetico che riporta le conclusione principali in lingua italiana e inglese. I verbali vengono fatti circolare tra i membri del GEV e approvati tramite email o utilizzando l'ambiente *software* predisposto dal CINECA.

4. Strumenti di valutazione

Il GEV sarà responsabile della valutazione finale di tutti i prodotti di ricerca assegnati, ed utilizzerà a tale scopo sia l'analisi bibliometrica sia la *peer review*. Il GEV valuterà tutti gli articoli su rivista con l'analisi bibliometrica, ed almeno il 10% degli stessi articoli anche con il metodo della *peer review*. Gli articoli su rivista da valutare in *peer review* saranno scelti mediante un campione casuale stratificato per sub-GEV. La scelta del campione terrà comunque conto della richiesta specifica di *peer review* segnalata tramite il modulo del CINECA per prodotti altamente specialistici e multidisciplinari. Dopo aver completato questa fase, il GEV studierà la concordanza tra analisi bibliometrica e *peer review* all'interno di ciascun sub-GEV e deciderà se ampliare la valutazione *peer review* degli articoli su rivista.

Per i prodotti di ricerca diversi da articoli su rivista (monografie, capitoli di libro, atti di congresso, ecc.), i Gruppi di Consenso (cfr. Sezione 4.1.2) effettueranno una prima valutazione interna al GEV basata sui seguenti criteri: rilevanza, originalità, internazionalizzazione, tenendo anche conto della diffusione della pubblicazione (internazionale, nazionale, locale), del prestigio accademico dell'editore e della collana specifica, della presenza di procedure editoriali aperte e trasparenti e della presenza di citazioni e recensioni in riviste internazionali. Se il Gruppo di Consenso unanimemente decide che il prodotto di ricerca sia "non valutabile" (perché appartiene a tipologie di prodotto escluse dalla VQR) oppure chiaramente "limitato" (perché si colloca nettamente al di sotto del 50% della scala di valore condivisa dalla comunità internazionale"), assegnerà il prodotto

di ricerca, rispettivamente, alla classe "Non valutabile" oppure "Limitato". In tutti gli altri casi il prodotto di ricerca sarà inviato alla *peer review*.

4.1. La valutazione peer review

4.1.1. Revisori esterni

Ciascun prodotto di ricerca da valutare in *peer review* sarà assegnato a due diversi revisori (normalmente, entrambi esterni al *panel*), scelti indipendentemente da due diversi componenti dei sub-GEV (o, se necessario, da due diversi componenti del GEV). In tutti i casi, almeno un revisore dovrà essere esterno al *panel*.

La scelta dei revisori eviterà conflitti di interesse tra revisori ed autori e tra revisori e sede di ricerca degli autori. L'indipendenza dei revisori sarà garantita prestando attenzione alla sede di ricerca dei revisori, alle collaborazioni di ricerca e, ove possibile, alla nazionalità.

Per evitare conflitti di interesse, il GEV adotterà le regole previste dalla VQR (Linee Generali per i Gruppi di Esperti della VQR, Sezione 5). In particolare, il Coordinatore assegnerà a revisori esterni i prodotti di ricerca dei componenti del GEV (se si tratta di monografie, capitoli su libro, atti di congresso, o se fanno parte del campione casuale degli articoli su rivista da valutare in *peer review*). ANVUR assegnerà a revisori esterni i prodotti di ricerca del Coordinatore del GEV.

Il GEV intende coinvolgere revisori esterni con un profilo di ricerca internazionale, un curriculum di altro profilo, testimoniato, in particolare negli ultimi anni, da un elevato numero di pubblicazioni nelle sedi di riferimento della comunità scientifica internazionale del settore, un significativo numero di citazioni e la necessaria competenza nella specifica area di valutazione. Il GEV preparerà un elenco di revisori esterni, stabilendo standard minimi di qualità scientifica, di impatto sulla comunità scientifica internazionale e di esperienza nella valutazione. Per minimizzare i conflitti di interesse, il GEV intende utilizzare ampiamente revisori che operano in Università ed istituzioni straniere.

Grande attenzione verrà posta al mantenimento dell'anonimato dei revisori, sia nella fase di predisposizione dell'elenco dei revisori, sia nella fase operativa di valutazione.

Per quanto attiene alla prima, il Presidente GEV consulterà la lista di revisori della propria area resa disponibile dal CINECA, e chiederà ai componenti GEV, tramite i coordinatori dei sub-GEV, di proporre un numero significativo di esperti che soddisfano ai criteri indicati nel paragrafo precedente.

Il Presidente GEV raccoglierà le indicazioni corredate di informazioni fornite sulla base di una scheda condivisa, e, anche con l'ausilio dei coordinatori di sub-GEV, provvederà a modificare la lista CINECA con integrazioni e/o cancellazioni.

Il processo di integrazione della lista continuerà per tutta la durata della valutazione, sulla base delle necessità che dovessero emergere a valle della trasmissione dei prodotti da parte delle strutture.

4.1.2. Valutazione di sintesi dell'analisi *peer review*

Ciascun revisore esterno fornirà una valutazione qualitativa sui prodotti di ricerca utilizzando la scheda predisposta per il revisore. La valutazione dei revisori *peer* si baserà su una apposita scheda revisore predisposta dal GEV, costituita da una serie di domande a risposta multipla e da un campo libero con numero limitato di parole. Il GEV trasformerà le indicazioni contenute nella scheda revisore in una delle 4 classi finali di merito. Per stabilire la valutazione di sintesi il GEV costituirà "Gruppi di Consenso", che saranno composti dai due componenti del GEV che hanno inizialmente assegnato il prodotto ai revisori esterni. I Gruppi di Consenso forniranno una valutazione di sintesi del prodotto di ricerca con il metodo della *informed peer review*, tenendo conto della valutazione dei due revisori esterni, degli indicatori disponibili sulla qualità e rilevanza del prodotto di ricerca e delle competenze del Gruppo di Consenso.

4.1.3. Conflitto di valutazioni

Se si verifica una divergenza tra valutazione *peer review* e valutazione bibliometrica, il Gruppo di Consenso fornirà una valutazione di sintesi. Per giungere ad una sintesi il Gruppo di Consenso considererà entrambi gli indicatori e potrà anche raccogliere ulteriori informazioni sulla qualità del prodotto di ricerca utilizzando il parere di un terzo revisore.

In caso di conflitto di valutazioni tra i componenti del Gruppo di Consenso, il Gruppo di Consenso sarà integrato con il Coordinatore del SUB-GEV o con il Coordinatore del GEV. Il Gruppo di Consenso potrà anche in questo caso decidere di utilizzare il parere di un terzo revisore esterno prima di concordare la valutazione di sintesi.

In tutti i casi, l'intero GEV sarà responsabile delle valutazioni finali, che saranno basate sulla cosiddetta *informed peer review*, e quindi terranno conto di tutti gli strumenti di valutazione disponibili (*peer review*, indicatori disponibili sulla qualità e rilevanza del prodotto di ricerca, competenza dei componenti del GEV).

4.2. Analisi bibliometrica degli articoli su rivista

4.2.1.Elenco delle riviste

Il GEV preparerà un elenco di riviste per ciascun sub-GEV. L'elenco iniziale si baserà sull'elenco Web of Science di Thomson Reuters (d'ora in poi WoS) e includerà tutte le riviste in WoS rilevanti per le aree di ricerca del GEV. Poiché la stessa rivista potrebbe apparire negli elenchi di due o più sub-GEV, per evitare duplicazioni ciascuna rivista sarà abbinata al sub-GEV più rilevante.

L'elenco iniziale sarà ampliato per includere riviste "non WoS" con elevato impatto in Google Scholar (orientativamente saranno aggiunte all'elenco di ciascun sub-GEV il 20-30% delle riviste). Le riviste saranno individuate utilizzando l'elenco di riviste U-Gov dell'Area 13 "Economia e Scienze Sociali" fornito dal CINECA. L'elenco include quindi ogni altro possibile elenco di riviste in cui hanno pubblicato i ricercatori italiani dell'Area 13 nel 2004-2010 (come Scopus, Scimago, ERA, Econlit, Repec, ecc). Le riviste che chiaramente non appartengono alle aree di ricerca coperte dal GEV saranno escluse e valutate da altri GEV o in collaborazione con altri GEV.

Le riviste aggiunte alla lista WoS dall'elenco U-Gov saranno individuate utilizzando due indicatori bibliometrici disponibili in Google Scholar: h-index della rivista nel 2004-2010 e numero di citazioni della rivista nel 2004-2010 in rapporto al numero degli articoli pubblicati nel 2004-2010 (o

periodi più brevi, se non disponibile). Aggiornamenti dell'elenco delle riviste sono possibili durante il processo di valutazione, in presenza di motivi validi e con il consenso della maggioranza del GEV.

Il GEV imputerà gli indicatori bibliometrici utilizzati per le riviste WoS alle riviste "non WoS" utilizzando gli indicatori bibliometrici disponibili in Google Scholar per entrambi i tipi di riviste. Le imputazioni saranno effettuate separatamente per gli elenchi di riviste di ciascun sub-GEV. Gli elenchi delle riviste WoS e di riviste "non WoS", e i rispettivi indicatori bibliometrici, saranno uniti in un unico elenco per ciascun SUB-GEV.

4.2.2. Indicatori bibliometrici

La classifica delle riviste sarà stilata separatamente per ciascun elenco di riviste per ogni sub-GEV (e possibilmente per una classificazione più fine delle aree di ricerca all'interno di ciascun sub-GEV) sulla base della distribuzione dell'Impact Factor (IF) delle riviste, dell'Impact Factor a cinque anni (5IF, se disponibile), o entrambi. Dato che IF e 5IF possono variare nel tempo, il GEV potrebbe decidere di utilizzare le medie per il periodo 2004-2010 (per periodi più brevi se non disponibile).

Il GEV è consapevole dei limiti di IF e 5IF, in particolare per il fatto che essi includono le autocitazioni. Tuttavia, il GEV ritiene che questi limiti non impediscono di classificare le riviste in ampi gruppi di merito. In ogni caso, il GEV intende verificare la robustezza della classifica delle riviste utilizzando indicatori alternativi (ad esempio, Article Influence Score).

Per verificare la robustezza della classificazione delle riviste il GEV intende anche confrontare l'elenco e le classifiche delle riviste con i dati Scopus. Infine, il GEV intende studiare la correlazione tra le classifiche delle riviste basate sugli indicatori bibliometrici disponibili nell'elenco ampliato WoS e in Scopus, per il sottoinsieme di riviste disponibile nei due elenchi, rivedendo le sue procedure nel caso di differenze significative nell'elenco delle riviste o nelle classifiche.

Il GEV classificherà le riviste, separatamente per ciascun elenco dei sub-GEV, in quattro classi di merito (Eccellente, Buono, Accettabile e Limitato), basandosi sui percentili della distribuzione di uno o più indicatori bibliometrici. Nel caso in cui le classi di merito "Eccellente", "Buono" o "Accettabile" congiuntamente non includano un numero minimo di riviste italiane (orientativamente 5 per ciascun sub-GEV), l'elenco di riviste di ciascun sub-GEV sarà ampliato per includere le migliori riviste italiane individuate sulla base degli indicatori bibliometrici disponibili in Google Scholar (h-index della rivista e numero di citazioni nel periodo 2004-2010), fino a quando il numero minimo di riviste italiane sia raggiunto. Le riviste italiane aggiunte saranno classificate in fascia "Accettabile". Nessun ampliamento dell'elenco è previsto se esso già include il numero minimo di riviste classificate almeno in fascia "Accettabile" per ciascun sub-GEV.

La classe di merito finale di ciascun articolo dipenderà anche dalle citazioni ricevute in WoS, ma solo per gli articoli con un "numero significativo di citazioni". Pertanto, se un articolo ha ricevuto un "numero significativo di citazioni nelle riviste WoS nel 2004-2010 in rapporto agli anni trascorsi dalla pubblicazione", sarà "promosso" di una classe (da Limitato ad Accettabile, da Accettabile a Buono, o da Buono ad Eccellente). A causa della mancanza di dati affidabili per l'analisi delle citazioni dei singoli articoli, nessuna correzione sarà applicata per gli articoli pubblicati in riviste

"non WoS". Nessuna "retrocessione" sarà applicata alla classe iniziale di merito di un prodotto di ricerca.

La classe di merito finale degli articoli nei gruppi Eccellente, Buono, Accettabile e Limitato dipenderà quindi dalla classe di merito iniziale della rivista (calcolata sulla base di indicatori bibliometrici) e dal numero di citazioni ricevute da ciascun articolo. Gli indicatori bibliometrici utilizzati e le soglie che identificano gli articoli con un "numero significativo di citazioni" saranno determinati nel corso del processo di valutazione, dopo aver verificato l'affidabilità e la solidità degli indicatori disponibili.

5. Conflitti di interesse

Il GEV adotterà le regole della VQR (Sezione 5 delle Linee Generali per i Gruppi di Esperti della VQR) per risolvere i conflitti di interesse tra i componenti del GEV e gli autori dei prodotti di ricerca.
In particolare, i membri del GEV si asterranno dal valutare o dall'assegnare ad altri membri del GEV o a esperti esterni:

a. prodotti di cui siano autori o co-autori;
b. prodotti di cui siano autori o co-autori congiunti, parenti o affini fino al 4° grado;
c. prodotti presentati da università presso cui i membri stessi abbiano o abbiano avuto un rapporto di lavoro o con le quali abbiano svolto incarichi o collaborazioni ufficiali negli anni a partire dal 1/1/2007;
d. prodotti presentati da enti di ricerca vigilati dal MIUR e da altri soggetti pubblici e privati sottoposti volontariamente alla VQR presso cui i membri stessi abbiano o abbiano avuto un rapporto di lavoro o con le quali abbiano svolto incarichi o collaborazioni ufficiali, inclusa l'affiliazione a enti di ricerca, negli anni a partire dal 1/1/2007.

Nei casi di cui al punto d) precedente, esiste conflitto di interesse:

i. nel caso in cui la struttura abbia una permanente strutturazione interna di tipo territoriale o disciplinare (es. sezione locale di ente di ricerca, istituto, dipartimento), limitatamente ai prodotti presentati dalla stessa articolazione;
ii. nel caso in cui la struttura non abbia una permanente strutturazione interna di tipo territoriale o disciplinare (es. sezione locale di ente di ricerca, istituto, dipartimento), in riferimento a tutti i prodotti presentati nei limiti in cui ciò sia possibile senza precludere la possibilità di valutare il prodotto;
iii. nel caso in cui la strutturazione interna abbia luogo a più livelli gerarchici (es. più istituti riuniti sotto un dipartimento) il conflitto di interesse sorge al livello più basso (es. membri del GEV affiliati ad istituti diversi di uno stesso dipartimento, sono in conflitto di interesse soltanto rispetto a prodotti presentati da autori appartenenti allo stesso istituto).

Nei casi di conflitto di interesse, il Presidente del GEV incaricherà delle procedure di valutazione un altro membro del GEV per i quali non vi siano conflitti di interesse. Nel caso di conflitti di interesse che coinvolgano il Presidente del GEV, l'assegnazione dei prodotti relativi sarà fatta dal coordinatore della VQR o da persona da lui incaricata.

National Agency for the Evaluation of Universities and Research Institutes
anvur
Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca

Evaluation of Research Quality
vQr
Valutazione Qualità della Ricerca

Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Appendice C

Classification, Imputation of Missing Values and Ranking of Journals

GEV 13 – Economics and Statistics

Classification, Imputation and Ranking of Journals

This document describes the construction of the database of journals of GEV 13 (Economics and Statistics), the method used to impute missing values of the bibliometrici indicators, and the journal ranking in groups A, B, C, D. The document is organized in five parts. Section 1 describes the construction of the database. Section 2 reports the definition of variables contained in the database. Section 3 presents descriptive statistics. Section 4 confronts the imputation of missing values, describing a simple two-step procedure and a more elaborate procedure based on multiple imputations. Section 5 presents the journal ranking. Section 6 discusses the ranking of Italian journals. Section 7 presents summary statistics on the distribution of journals in the different merit classes by research areas. An Appendix contains the technical aspects of the multiple imputation procedure.

1. The journal list

The initial list has been based on Thomson Reuters Web of Science (from now on WoS) and includes all journals in ISI-JCR Social Science Edition belonging to the subject categories relevant to GEV 13, namely:

DI (Business)
DK (Business, Finance)
FU (Demography)
GY (Economics)
NM (Industrial Relations and Labor)
PS (Social Sciences, Mathematical Methods)
PE (Operations Research and Management Science)
XY (Statistics and Probability).

The initial list also includes many journals in ISI-JCR Science Edition belonging to subject categories relevant to GEV 13:

AF (Agricultural Economics)
JB (Environmental studies)
KU (Geography)
NE (Public, Environmental and Occupational Health)
PO (Mathematics, Interdisciplinary Applications)
WY (Social Work)
YQ (Transportation).

The initial list has been expanded using the U-Gov list of journals for Area 13 “Economics and Social Sciences” provided by CINECA, in which at least one Italian researcher of Area 13 has published in 2004-2010.

From the U-Gov dataset the following journals have been excluded:

1) journals that clearly fall outside the scientific research areas covered by GEV 13.
2) working paper series, collections/reports of Department/Faculty/ Research Institution that do not satisfy the rules described in the ANVUR website (FAQ Section);
3) journals for which the bibliometric indicator (h-index) is missing in Google Scholar for the period 2004-2010 (or shorter periods for recent journals);
4) journals for which the bibliometric indicator (h-index) is negligible in Google-Scholar (GS) The rule has been to exclude journals with h-index lower than 3 in 2004-2010.
5) Journals that are too recent to have reliable indicators in Google-Scholar. Prominent examples are the American Economic Journals (macroeconomics, microeconomics; applied economics; economic policy) and the Annual Review of Economics.

Articles published in journals that are not classified will be evaluated in peer review by GEV 13, or – in case of multidisciplinary journals - in collaboration with other GEV. Each journal has then been assigned to one of 5 areas:

A - Business, Management and Finance;
E - Economics;
H - Economic History and History of Economic Thought;
S - Statistics and Applied Mathematics;
G - General journals: Science, Nature, PNAS.

Each journal has been matched to one area (without duplications) to avoid different rankings in different areas.

In March and April 2012 GEV members were consulted several times about the database. In April Scientific societies (SIS, SIE, AIDEA, SIdE, SISE, AMASES) were consulted and supplied very useful suggestions. The first version of the journal list, published on ANVUR website on April 30th , classified journals in three areas: A, E, S. The statement accompanying the journal list specified that:

- GEV reserves the possibility of using narrower journal classification areas (to allow bibliometric comparisons in homogeneous sets of journals);
- the list can contain misclassification and measurement errors of the bibliometric indicators. Hence it remains at the disposal of GEV and the scientific community for following checks and integrations.

On May 10th, GEV announced the creation of a new classification area (History: H). The decision was motivated by the large differences in bibliometric indicators of journals of Economic History and History of Economic Thought (less than half compared to the journals of other areas). In May GEV received further comments and suggestions from the scientific community – scientific societies and individual researchers – concerning the following issues:

- misclassification in the four areas (A, E, S, H);
- presence of journals in WoS;

- values of the h-index;
- proposal to include journals that meet the GEV classification requirements;
- proposals to exclude journals published after 2008;
- exclusion of journals pertaining to other disciplines;
- errors in the name or ISSN of journals.

GEV analyzed each of the comments and incorporated them in the updated version of the list, which was posted on the ANVUR Website on the 12th June 2012. Between June 12 and July 11 GEV received few more comments, and approved a first version of the journal list on July 11. Between July 12 and August 20, GEV received further comments on the journal list and ranking of journals, implementing few more changes.[2] Overall, the journal list has been exposed for more than 3 months to comments and suggestions from the scientific community.

2. Definition of variables

In this section we describe the definition of the variables used to classify journals, and contained in our dataset.

Impact Factor (henceforth IF): is a measure, provided by JCR, of the frequency with which the "average article" in a journal has been cited in a particular year or period. The annual impact factor of a journal is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years. In our dataset we report the IF released in December 2010.

5-year Impact Factor (henceforth IF5): is defined as the ratio between citations and citable items published in 5 years period. In our dataset we report the IF5 released in December 2010.

Article Influence Score (henceforth AIS): it is also provided by JCR and determines the average influence of a journal's articles over the first five years after publication. It is calculated by dividing a journal's Eigenfactor Score by the number of articles in the journal, normalized as a fraction of all articles in all publications. The Eigenfactor Score calculation is based on the number of times articles from the journal published in the past five years have been cited in the JCR year, but it also considers which journals have contributed these citations so that highly cited journals will influence the network more than lesser cited journals. References from one article in a journal to another article from the same journal are removed, so that Eigenfactor Scores are not influenced by journal self-citation. This measure is roughly analogous to IF5 because it is the ratio of the number of journal's citations to number of articles it has published over a period of five years. The mean value of AIS is 1.00. A score greater than 1.00 indicates that an article in the journal has above-average influence A score less than 1.00 indicates that an article in the journal has below-average influence. In our dataset we report the AIS released in December 2010.

[2] Changes are detailed and motivated in the Anvur website.

h-index: A journal has index *h* if *h* of its *N* papers have at least *h* citations each, and the other (*N*-*h*) have no more than *h* citations each. We computed the h-index in Google Scholar in 2004-2010. Data were collected in April 2012, and checked throughout May 2012.

Italian journals: a journal is classified as "Italian" if (1) it publishes papers in Italian; or (2) it publishes papers in English (part or all) and is printed and edited in Italy; or(3) is printed by international publishers on behalf of Italian institutions or associations.

English journals: of the residuals journals ("non-Italian") a journal is classified as "English" if (1) it publishes papers in English; or (2) it is printed and edited in the UK, USA, Australia.

Other language: of the residuals journals ("non-Italian") a journal is classified as "Other" if it publishes papers in languages other than Italian and English, or if it is printed and edited in countries other than those previously mentioned.

ISI journal: the journal is present in the WoS database.

3. Descriptive statistics

The overall list of journals includes 1,906 journals. Excluding the 3 journals classified in G, the list contains 1903 journals in the 4 areas:

A - Business, Management and Finance: 767 journals (40%);
E - Economics: 643 journals (34%);
H - Economic History and History of Economic Thought: 48 journals (2%);
S - Statistics and Applied Mathematics: 445 journals (23%).

Overall, ISI Journals represent 49% of the list. The fraction of ISI journals varies by area, ranging from 40% for H to 42% for A, 52% for E and 56% for S (Table 1.1). Italian Journals represent 5.7% of the total number of journals. In particular in areas A and S only 3% of journals are Italian while in E and H the number of Italian journals is higher (9.6% and 12.5%, respectively) (Table 1.2).

Table 2.1 reports basic statistics for IF by area: mean, standard deviation, 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles, inter-quartile range. The IF index is available for 912 ISI journals, with a mean of 1.19 and a standard deviation of 0.97. The average IF varies among areas, and is highest for A (1.47) and lowest for H (0.49).

The 5-year Impact Factor (IF5) and Article Influence Score (AIS) are available for a subset of ISI journals (648). Averages and percentile data by areas of these two indexes show the differences in citation patterns among areas and the lowest data for history journals (Table 2.2 and 2.3). It is apparent that (except for area H, where large differences persist) the distribution of AIS is more similar across areas than the distribution of IF5. This is true for each of the percentiles reported in the table, as well as for the inter-quartile range.

The h-index collected in Google Scholar also reveals differences in citations standards across areas: the lowest mean h-index is once again in area H, while the highest is in area A (Table 2.4).

Correlation coefficients are reported in logarithms because the imputation uses the logarithms of the variables in order to make their distribution closer to the normal one and to reduce heteroskedasticity (see Section 4). The correlation between the three bibliometric indicators available in WoS is, not surprisingly, high: for instance, the correlation between log(IF) and log(IF5) is above 0.9 in all areas; the correlation between log(IF5) and log(AIS) is higher than 0.8 in all areas.

The h-index is strongly and positively correlated with each of the three bibliometric indicators available in WoS (IF, IF5, AIS). In particular:

- in areas A and E the correlation between log(IF5) and log(AIS) with log(h) exceeds 0.7;
- in area H the correlation ranges from 0.61 (for IF) to 0.72 (for IF5);
- in area S the correlation ranges from 0.65 (for AIS) to 0.73 (for IF5).

Such values make us confident that imputation of missing values of IF, IF5 and AIS will be based on a strong predictor of each of the three variables.

4. Imputation of bibliometric indicators

In this section we describe the procedures used to impute the missing values of the three bibliometric indicators of interest, namely IF, IF5 and AIS. However, given that IF is subject to high year-to-year variability due to the short time period (two years) used for its computation, we will devote most of our attention to IF5 and AIS.

The prevalence of missing values for all three indicators in each of the four scientific areas is shown in Table 4.1. In column (1) the total number of journals is recorded, while in columns (2) and (3) one can see the raw number of journals with missing values for IF and the percentage of those journals, respectively. In columns (4) and (5) the same calculations are performed for IF5 and AIS (these two indicators have identical patterns of missingness, as the AIS can be defined only when IF5 is defined as well). We note the prevalence of missing values is notable for both IF and IF5/AIS. In fact, in the latter case the prevalence of missing values is larger by about 11-15%. When looking at particular areas, it is clear that journals belonging to areas H and A are the most affected by the problem of missing values, while journals in area S are the least affected.

Before proceeding with the imputation of missing values, it is useful to inspect the distribution of the non-missing values of the bibliometric indicators, as this distribution could affect the choice of imputation model. The kernel densities of the levels of IF5 and AIS are shown in Fig 4.1.A, and it is immediately clear that their distributions are very asymmetric. In particular, they are skewed to the right with substantial right tails, and this happens in all four areas. This fact is corroborated by the calculation of the skewness and kurtosis of these two indicators, displayed in Table 4.2, columns (1)-(4). We note that in all four areas the distributions of both indicators are very different from a normal distribution, which has a skewness equal to zero and a kurtosis equal to three. The worst

affected areas are E and A, while the least affected is H (possibly because the sample size is small for that area).

Such large skewness and kurtosis of the distributions of the three bibliometric indicators makes the estimation of regression models in the levels of the indicators quite problematic, as the resulting estimates are extremely likely to be unduly influenced by outliers found in the long right tail of these distributions. Therefore, we chose to estimate our models after taking the logarithms of the values of the indicators. The logarithmic transformation (which does not change the ranking of values as it is a strictly increasing transformation) makes the distribution of the non-missing values much more symmetric and closer to the normal distribution, as can be clearly seen in Fig. 1.B, and in the much improved skewness and kurtosis values of the logarithms shown in columns (5)-(8) of Table 4.2. In areas E, A, and S, and for both IF5 and AIS, the skewness is closer to zero and the kurtosis closer to four after the logarithmic transformation, and thus closer to the values of the skewness and kurtosis of the normal distribution, which is what we wanted to achieve.

Having thus decided to perform our estimation in logarithms, we chose the following two imputation models:

i) A baseline imputation model (henceforth BIM), in which the logarithm of each of the three bibliometric indicators (IF, IF5, AIS) is regressed on the logarithm of the h-index (the level of which is skewed to the right before the logarithmic transformation) and a constant.[3] We can use the h-index as a predictor because it is never missing. The estimation is performed separately in each area and the estimation sample for each indicator/area combination consists always of the observations with non-missing values for the particular indicator in the particular area. After the regression is performed, we assign to each observation with a missing value the mean prediction from the regression, which will thus be our imputed value for that particular observation.

ii) A more elaborate multiple imputation model (henceforth MIM) in which instead of trying to come up with a single imputed value for each missing one, we try instead to trace the distribution of the missing value by creating multiple imputed values for each observation. The principle of multiple imputation was first elaborated by Rubin (see, e.g, Rubin, 1987), and is currently widely used in micro data surveys.

The advantage of MIM over BIM is that tracing the distribution of each imputed value should better capture the uncertainty that accompanies the imputation process. Instead of getting a point prediction as in BIM, by using MIM we recognize that our imputed value is characterized by uncertainty. Therefore, it is a good idea to produce a range of imputed values for each missing one. This should allow us to get better estimates not only of the mean of the missing value but also of the variance due the imputation process. This is important because overlooking this uncertainty typically results in biased standard errors of any statistical estimates derived from the sample after imputation. On the other hand, a single imputed value as produced by BIM does not allow this uncertainty to show up anywhere and is effectively treated the same way as a non-missing value in any statistical procedure.

[3] Standard errors are computed using the “robust” option in Stata.

In the MIM each variable to be imputed is regressed not only on any non-missing predictors (like the logarithm of the h-index) but also on the other variables to be imputed The estimation sample consists once more of the observations with non-missing values for the particular indicator in a particular area, but now the predictors can have imputed values. For example, for the imputation of AIS we will use as predictors IF and IF5. The latter variable can have imputed values in the sample of non-missing observations for AIS; given the high correlation of AIS with IF and IF5, including the two indicators should increase the predictive power of the regression model.[4]

The MIM process runs iteratively, and the number of iterations is determined by whether the distribution of imputed values shows convergence, which has a particular meaning in a MIM context: convergence obtains when the variance across implicates becomes small enough compared to the variance within implicates, i.e., when the predicted values do not differ too much across implicates. We currently use 100 iterations and check for this convergence, and use the data from the last iteration as our final imputations.

The full list of regressors includes the logarithms of h and the remaining two indicators in each case (e.g. IF5 and AIS when imputing IF), as well as their squares. We also include an indicator for whether a journal is published in English, as this information can be found in our database. Even though the regression is performed in logarithms, we exclude from our estimation sample observations with values of the dependent variable lower or equal to the 1st percentile and higher or equal to the 99th percentile in order to further diminish the influence of outliers on our estimates. As a result, the MIM estimation samples are slightly smaller (by about six observations) than the BIM ones.

As already discussed, we produce multiple imputations for each missing value and record each imputed value in a separate dataset (called implicate). We use 500 implicates in the current implementation of the MIM, i.e., we produce 500 imputed values for each missing one. Our final imputed value of the logarithm of each indicator for a particular observation will be the average over the 500 imputed values for that observation, as discussed in Rubin (1987). Given that the estimation sample for area H is so small, we do not use the MIM method in that area.

We graph the regression line from the BIM model in Figure 4.2, and we can see that for all areas the fit of the regression line for both IF5 and AIS is reasonably good. The full estimation results for both BIM and MIM are shown in Tables 4.3.A and 4.3.B for IF5 and AIS, respectively.[5] We note that the adjusted R^2 of the BIM model is quite high, which indicates a good predictive power of the model, especially if one takes into account the fact that only one regressor, namely the logarithm of the h-index, is used.

As already discussed, the MIM includes a richer set of regressors than BIM. In particular, this set includes the two remaining indicators in each case, which are highly correlated with the indicator to be imputed (e.g., IF and IF5 are very good predictors of AIS). As a result, the adjusted R^2 (a

[4] The particular implementation of the MIM model that we are using is due to van Buuren et al. (2006). The details of our MIM model can be found in Appendix A

[5] The estimates of the MIM are taken from the final 100th iteration and the first implicate dataset. Estimates from the other implicate datasets in the 100th iteration are very similar to the ones shown here.

measure of it that penalizes the increased number of predictors) of the MIM is quite higher than that of the BIM.

In Fig. 4.3 we can see the distributions of the imputed values (in logarithms) for ten different implicates in the final (100^{th}) iteration of the MIM model for the case of AIS in area E. For comparison, we also graph the distribution of the non-missing values. We can see that the distribution of imputed values differs across implicates, but not exceedingly so. The same is true in almost all indicator/area combinations. Furthermore, in all 500 implicates (i.e., not just those shown in Fig. 4.3) the distribution of imputed values lies clearly to the left of the distribution of non-missing values, which implies that the imputed values of the logarithm of AIS in area E are typically smaller than the non-imputed ones. This is to be expected, as journals with missing information for AIS have typically lower values of the h-index and the IF (when non-missing), and both these variables are predictors in the regression equation for AIS. However, the distribution of the non-imputed values has a typically fatter left tail than that of the imputed values, which implies that there are quite a few low non-missing values. This in turn should make our regression estimates more robust, as a wide range of values of the dependent variable is used in the estimation.

Importantly, the fact that the distribution of imputed values lies to the left of that of non-imputed ones for both BIM and MIM is true for all indicator/area combinations.

It is also interesting to compare the distributions of the bibliometric indicators across imputation methods. When we graph these distributions in Fig. 4.4 (using the whole sample) we note that MIM tends to produce a thinner left tail than BIM.[6] Otherwise, in both imputation models the main mass of the distribution lies in roughly the same area for all area/indicator combinations.

[6] There is no graph for area H because MIM is not used in that case.

5. Ranking of journals

Having produced the imputations using both BIM and MIM we proceed to compare them in a more formal way by examining the differences of the journal rankings that they imply. In order to rank journals we first create deciles of the distributions of the logarithms of IF5, AIS and h in each area. These deciles are computed using both the non-imputed and the imputed values. We then concatenate the deciles into four levels using the following rule: the lowest five deciles are put into level one, the sixth decile into level two, the seventh and eighth deciles into level three and the top two deciles into level four. Having thus created the four levels, we compare how the classification of journals in these levels differs between BIM and MIM for the ranking according to each bibliometric indicator.

5.1. Comparison between BIM and MIM

The results of this comparison can be seen in Table 4.4, columns (1)-(5) for the ranking according to IF5, and columns (6)-(10) for the one according to AIS. Along the vertical dimension we have the levels defined using BIM, and along the horizontal dimension the levels defined using MIM. The numbers in red are in the diagonal of the cross-tabulation, i.e., they show the number of journals that are classified in the same level by both BIM and MIM. The cells that are farthest from the diagonal denote journals for which the two imputation methods give the most divergent results. We note that, overall, both methods give quite similar results in all area/indicator combinations (area H is not included because MIM is not used in this case).

The congruence of the level classification between BIM and MIM can be seen even more clearly in Table 4.5, which shows the divergence between the two imputation methods. This divergence can be assessed by comparing the difference between the BIM and MIM level ranking, using both IF5 and AIS. For example, we note that for AIS in area E there are 52 journals with a difference in ranking equal to minus one, i.e., they are ranked one level lower using BIM compared to using MIM. On the other hand, there are 28 journals in area S with a difference in ranking by IF5 equal to plus one, i.e. they are ranked higher by one level using BIM than using MIM.

In order to better judge the difference in rankings between the two methods, we report for each area/indicator combination the percentage of journals for which the difference in ranking is between minus one and plus one. Clearly, these are journals for which both imputation methods produce rankings that are quite similar. We note that the overwhelming majority of journals (on average 95%, with the lowest percentage being 92% for AIS in area A) have a ranking difference of at most one in absolute value. In fact, most journals have the exact same ranking.

These results make us conclude that, while BIM and MIM might give in some cases different results for individual journals, for the purposes of classifying journals into the aforementioned four levels, both methods give essentially equivalent results. We will thus use for our final journal classification the ranking produced by BIM, as it is the simpler and more easily implementable imputation method of the two.

5.2. Comparison between IF5, AIS and h-index

Having chosen BIM, we are now interested in looking at the differences in journal rankings across different indicators. To that effect, we compare rankings using the logarithms of IF5, AIS and h, and the cross-tabulations are shown in Table 4.6. Once more, we observe that most journals lie in the diagonal of every comparison, i.e. they have the same ranking using both indicators.

This pattern can be seen more clearly once more by looking at the distribution of the differences in ranking, shown in Table 4.7. We note once more that the overwhelming majority of journals are ranked very similarly by all indicators: the lowest percentage of journals with a ranking difference between minus one and plus one is 94.8% in area S when comparing rankings according to AIS and h. The percentage of journals with exactly the same ranking is typically a bit lower, but still very high: the lowest percentage of journals identically ranked is about 75% in area S when comparing rankings according to AIS and h-index. Once more these results are not surprising, as IF5, AIS and h-index are all strongly positively correlated, and the h-index is a crucial predictor when imputing both IF5 and AIS.

Given the strong correlation between AIS and IF5, in principle one could use either one for the final classification. The AIS excludes journal self-citations, and gives more weight to citations received from higher ranked journals. On the other hand, in very small and specialized research areas IF5 is not necessarily an indicator of "self-promotion" on the part of the journal, because highly specialized researchers tend to cite more often journals from their own field, and are less cited by researchers in different fields.

Given these considerations, in the meeting of July 11, 2012 GEV decided to use the maximum of the AIS and IF5 ranks as final rank. With respect to a situation in which one were to apply either the AIS or the IF5, the percentage of journals that switch classes is 6.7% on average (4% in area A, 7% in area E, 8% in area H, and 10% in area S). The number of switches in Area S is slightly larger given the higher disagreement of AIS and IF5 scores in this area. We speculate that such higher divergence is likely to be due to a higher degree of heterogeneity in journals and sub-fields belonging to area S compared to other fields.

6. Ranking of Italian journals

The last step of our classification of journals is to consider Italian journals. GEV evaluation criteria state that the journal list for each area will be expanded to include the best Italian journals assessed on the basis of the bibliometric indicators, and that the Italian journals added will be classified in group C.

By "Italian journals" GEV defines journals that publish only Italian papers, a mix of papers in Italian and other languages, journals published by Italian publishers, and journals published by international publishers on behalf of Italian institutions or associations. The list of Italian journals is

displayed in Table 6.1. The table shows that only three Italian journals are ranked higher than Level 1 (corresponding to VQR Class D) according to either the AIS or IF5 criteria. All other journals are ranked in Class D because articles published in them are cited relatively infrequently. Note that all Italian journals have imputed values for AIS and IF5; therefore, their ranking is effectively based on the h-index reported in the table.

On the basis of the rankings in Table 6.1, in the meeting of July 11, 2012 GEV adopted the following rules for the number of upgrades: (1) a positive number in each area; (2) selection according to bibliometric indicators; (3) an average of 5 journals for each area. The total number of selected journals was 20 and the selected journals are displayed in Table 6.1. The cut-offs to upgrade journals from D to C vary slightly by area, being lowest in area A and H, and highest in area S and E; note that in no case the h-index cut-offs are less than 6.

Summary statistics for the final ranking of Italian journals are reported in Table 6.2. The table shows that in Area A 24% of Italian journals are classified as “Level 2” (corresponding to the VQR Class C). In Area E the fraction of journals classified as Levels 2 or 3 is 18%; in Area H and S 33% and 25% of journals, respectively, are classified as Level 2.

7. Final classification of journals

Table 7 tabulates the final distribution of journals by area. Overall, 48.7% of journals are Level 1 (VQR Class D), 9.4% Level 2 (VQR Class C), 18.5% Level 3 (VQR Class B), and 23.4% Level 4 (Class A). The slightly different proportions compared to the VQR suggested guidelines (50/10/20/20) reflect the rule of the maximum between AIS and IF5 ranks, the presence of ties in the imputed values of AIS and IF5 and the decision to upgrade some Italian journals to Level 2. In Level 1 (VQR Class D), the fraction of papers is similar in all research areas.. In Level 4 (VQR Class A) the fraction is slightly higher then average in Area S (25.2%), and slightly below average in Areas E and H (22.4% and 20.8%, respectively). In terms of absolute numbers, the largest number of journals ranked in Level 4 is Area A (172 journals), followed by E (152 journals), S (112 journals) and H (10 journals).

The final classification of each journal article will also depend on citations received in WoS. Hence, if an article has received at least five citations per year in 2004-2010 (or sub-periods), it will be upgraded by one ranking. Given the lack of reliable data for citation analysis of individual journal articles, no correction will be applied to journals not classified in WoS. The final classification of articles therefore depends on the initial journal ranking (based on the combined IF5/AIS indicators) and the number of citations received by each article.

Finally, GEV plans to evaluate 10% of the articles using peer review. Journal articles evaluated in peer review will be chosen by a random sample stratified by research areas. After completing this process, GEV will study the congruence between bibliometric analysis and peer review within each area, and will then decide whether to expand peer review evaluation or not.

References

van Buuren, S., J.P.L. Brand, C.G.M. Groothuis-Oudshoorn, and D.B. Rubin. (2006). "Fully Conditional Specification in Multivariate Imputation." Journal of Statistical Computation and Simulation, 76:1049-64.

Christelis, D. (2011). "Imputation of Missing Data in Waves 1 and 2 of SHARE". CSEF Working Paper No. 278. Available at http://www.csef.it/WP/wp278.pdf.

Lepkowski, J. M., T. E. Raghunathan, J. Van Hoewyk, and P. Solenberger. (2001). "A Multivariate Technique for Multiply Imputing Missing Values using a Sequence of Regression Models." Survey Methodology, 27: 85-95.

Little, R. E. and D. B. Rubin. (2002). Statistical Analysis of Missing Data, 2nd Edition. New York, NY: John Wiley & Sons.

Rubin, D. B. (1987). Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons.

Schafer, J.L. (1997). Analysis of Incomplete Multivariate Data. Boca Raton, FL: Chapman and Hall.

Tanner, M.A. and W.H. Wong. (1987). "The Calculation of Posterior Distributions by Data Augmentation (with discussion). " Journal of the American Statistical Association, 82: 528-550.

Appendix. Detailed description of the MIM

The imputation methodology that we use is the fully conditional specification method (FCS) of van Buuren, Brand, Groothuis-Oudshoorn, and Rubin (2006, henceforth BBGR), and the exposition from this point on follows closely theirs.[7] Let $\boldsymbol{Y} = (\boldsymbol{Y_1}, \boldsymbol{Y_2}, \ldots \boldsymbol{Y_k})$ be a $\boldsymbol{n \times K}$ matrix of K variables (all potentially containing missing values) in a sample of size n. In our case $\boldsymbol{K}$ is equal to three, as we are imputing the logarithms of IF, IF5 and AIS. $\boldsymbol{Y}$ has a multivariate distribution characterized by a parameter vector $\boldsymbol{\theta}$, denoted by $\boldsymbol{P}(\boldsymbol{Y}; \boldsymbol{\theta})$. The objective of the imputation procedure is to generate imputed values for the missing part of $\boldsymbol{Y}$ (denoted by $\boldsymbol{Y_{mis}}$) that, combined with the non-missing part $\boldsymbol{Y_{obs}}$, will reconstitute as closely as possible the joint distribution $\boldsymbol{P}(\boldsymbol{Y}; \boldsymbol{\theta})$.

One way to proceed would be to assume a fully parametric multivariate density for $\boldsymbol{Y}$, and starting with some priors about $\boldsymbol{\theta}$ to generate imputations of $\boldsymbol{Y_{mis}}$ conditional on $\boldsymbol{Y_{obs}}$ (and on any other vector of variables $\boldsymbol{X}$ that are never missing, like the h-index in our case).

An alternative to specifying a joint multivariate density is to predict any given variable in $\boldsymbol{Y}$, say $\boldsymbol{Y_k}$, conditional on all remaining variables in the system (denoted by Y_{-k}) and a parameter vector $\boldsymbol{\theta_k}$. We apply this procedure to all $\boldsymbol{K}$ variables in $\boldsymbol{Y}$ in a sequential manner, and after the last variable in the sequence has been imputed then a single iteration of this process is considered to be completed. This way the K-dimensional problem of restoring the joint density of $\boldsymbol{Y}$ is broken into $\boldsymbol{K}$ one-dimensional problems of conditional prediction. This breakdown has two principal advantages over the joint approach:

- It can readily accommodate many different kinds of variables in $\boldsymbol{Y}$ (e.g. binary, categorical, and continuous). This heterogeneity would be very difficult to model with theoretical coherence using a joint distribution of $\boldsymbol{Y}$.
- It easily allows the imposition of various constraints on each variable (e.g. censoring), as well as constraints across variables.

The principal drawback of this method is that there is no guarantee that the $\boldsymbol{K}$ one-dimensional prediction problems lead to convergence to the joint density of $\boldsymbol{Y}$. Because of this potential problem, BBGR ran a number of simulation tests, often complicated by conditions that made imputation difficult, and found that the FCS method performed very well. Importantly, it generated estimates that were generally unbiased, and also good coverage of the nominal confidence intervals.

As the parameter vector $\boldsymbol{\theta}$ of the joint distribution of $\boldsymbol{Y}$ is replaced by the $\boldsymbol{K}$ different parameter vectors $\boldsymbol{\theta_k}$ of the $\boldsymbol{K}$ conditional specifications, BBGR propose to generate the posterior distribution of $\boldsymbol{\theta}$ by using a Gibbs sampler with data augmentation.

[7] The exposition in this Appendix is based on Christelis (2011).

Let us suppose that our imputation process has reached iteration *t*, and that we want to impute variable $\boldsymbol{Y_k}$. We first estimate a statistical model[8] with $\boldsymbol{Y_k}$ as the dependent variable (using only its observed values), and the variables in $\boldsymbol{Y\text{-}_k}$ as predictors. For every element of $\boldsymbol{Y\text{-}_k}$ that precedes $\boldsymbol{Y_k}$ in the sequence of variables, its values from iteration *t* are used (i.e., including the imputed ones). On the other hand, for every element of $\boldsymbol{Y\text{-}_k}$ that follows $\boldsymbol{Y_k}$ in the sequence, its values from iteration *t-1* are used.

After obtaining the parameter vector $\boldsymbol{\theta_k}$ from our estimation, we make a draw θ_k^* from its posterior distribution[9], i.e., we have

$$\boldsymbol{\theta_k^{*(t)} \sim P\left(\theta_k \left| Y_1^{(t)}, \ldots, Y_{k-1}^{(t)}, Y_{k,obs}, Y_{k+1}^{(t-1)}, \ldots, Y_K^{(t-1)}\right.\right)} \qquad (1)$$

The fact that only the observed values of $\boldsymbol{Y_k}$ are used in the estimation constitutes, as BBGR point out, a deviation from most Markov Chain Monte Carlo implementations, and it implies that the estimation sample used for the imputation of any given variable will include only the observations with non-missing values for that variable.

Having obtained the parameter draw $\theta_k^{*(t)}$ at iteration *t* we can use it, together with $Y_{-k}^{(t)}$ and the observed values of $\boldsymbol{Y_k}$, to make a draw from the conditional distribution of the missing values of Y_k. That is, we have

$$\boldsymbol{Y_k^{*(t)} \sim P\left(Y_{k,mis} \left| Y_1^{(t)}, \ldots, Y_{k-1}^{(t)}, Y_{k,obs}, Y_{k+1}^{(t-1)}, \ldots, Y_K^{(t-1)}; \theta_k^{*(t)}\right.\right)} \qquad (2)$$

As an example, let us assume that Y_k represents the logarithm of the value of a particular bibliometric indicator, and that we want to impute its missing values at iteration *t* via ordinary least squares, using the variables in $Y_{-k}^{(t)}$ as predictors. We perform the initial estimation, and obtain the parameter vector $\boldsymbol{\theta_k^{(t)} = \left(\beta_k^{(t)}, \sigma_k^{(t)}\right)}$, with $\boldsymbol{\beta_k^{(t)}}$ denoting the regression coefficients of $Y_{-k}^{(t)}$, and $\sigma_k^{(t)}$ the standard deviation of the error term. After redrawing the parameter vector $\theta_k^{*(t)}$ using (A.1), we first form a new prediction that is equal to $\boldsymbol{Y_{-k}^{(t)}\beta_k^{*(t)}}$. Then, the imputed value $\boldsymbol{Y_{k,i}^{*(t)}}$ for a particular observation *i* will be equal to $\boldsymbol{Y_{-k,t}^{(t)}\beta_k^{*(t)}}$ plus a draw of the error term (assumed to be normally distributed with a standard deviation equal to $\sigma_k^{*(t)}$ [10]). The error draw for each observation with a missing value for Y_k is made in such a way as to observe any bounds that have been already placed on the admissible values of Y_k for that particular observation. These bounds can have many sources, e.g. overall minima or maxima imposed for the particular variable.

The process described in (A.1) and (A.2) is applied sequentially to all K variables in *Y*, and after the imputation of the last variable in the sequence (i.e., Y_k) iteration *t* is considered complete. We thus end up with an example of a Gibbs sampler with data augmentation (Tanner and Wong, 1987) that

[8] In our case the statistical model is always a linear one, but in other cases nonlinear models can be used (e.g., probit, multinomial logit), depending on the nature of Y_k.
[9] The formulas used for redrawing the parameter vector can be found in Appendix A of BBGR.
[10] As already discussed in the text, the estimation of all models of amounts is done in logarithms in order to make our conditional specifications more compatible with the maintained assumption of normality.

produces the sequence {($\theta_1^{(t)}, \dots, \theta_K^{(t)}, Y_{mis}^{(t)}$): t=1,2,...}. The stationary distribution of this sequence is P($Y_{mis}, Y_{obs}; \theta$), provided that convergence of the imputation process is achieved.

As Schafer (1997) points out, a sufficient condition for the convergence to the stationary distribution is the convergence of the sequence {$\theta_1^{(t)}, \dots, \theta_K^{(t)}$} to the conditional distribution of the parameter vector P($\theta | \mathbf{Y}_{obs}$), or, equivalently, the convergence of the sequence {$Y_{mis}^{(t)}$} to the conditional distribution of the missing values $\boldsymbol{P(Y_{mis} \mid Y_{obs})}$. Hence, in order to achieve convergence to the stationary distribution of $\boldsymbol{Y}$, we iterate the Gibbs sampler till we have a number of iterations indicating convergence of the distributions of the missing values of all the variables in our system.

One important feature of the FCS method (shared with several other similar approaches found in the imputation literature[11]) is that it operates under the assumption that the missingness of each variable in $\boldsymbol{Y}$ depends only on other variables in the system and not on the values of the variable itself. This assumption, commonly known as the missing at random (MAR) assumption, is made in the vast majority of imputation procedures applied to micro datasets. It could be argued, however, that it is unlikely to hold for all variables: for example, missingness in AIS could depend on whether the journal might have a high or low citation count and thus high or low potential AIS. This would be a case of data missing not at random (MNAR), and, if true, would present major challenges for the construction of the imputation model.

Some evidence on the consequences of the violation of the MAR assumption comes from the results of one of the simulations run by BBGR, which exhibits a NMAR pattern. In addition, BBGR use in this simulation conditional models that are not compatible with a single joint distribution. Even in this rather pathological case, however, the FCS method performs reasonably well, and leads to less biased estimates than an analysis that uses only observations without any missing data. As a result, BBGR conclude that the FCS method (combined with multiple imputation) is a reasonably robust procedure, and that the worry about the incompatibility of the conditional specifications with a joint distribution might be overstated.

One further issue to be addressed is how the iteration process is started, given that, as described above, one needs in any given iteration to use imputed values from the previous iteration. In other words, we need to generate an initial iteration, which will constitute an initial condition that will provide the lagged imputed values to the first iteration. This initial iteration is generated by imputing the first variable in the system based only on variables that are never missing (namely the logarithm of the h-index and the English language indicator), then the second variable based on the first variable (including its imputed values) and the non-missing variables, and so on, till we have a complete set of values for this initial condition. Having obtained this initial set of fully imputed values, we can then start the imputation process using the already described procedures, as denoted in equations (1) and (2).

Once we have obtained the imputed values from the last iteration, we end up with five hundred imputed values for each missing one, i.e., with five hundred different complete datasets that differ from one another only with respect to the imputed values. We then need to consider how to use the five hundred implicate datasets in order to obtain estimates for any magnitude of interest (e.g. descriptive statistics or coefficients of a statistical model).

[11] A similar imputation procedure is proposed by Lepkowski, Raghunathan, Van Hoewyk, and Solenberger (2001). See also BBGR for references to a number of other approaches that have significant similarities to theirs.

Let m = 1,…., M index the implicate datasets (with M in our case equal to five hundred) and let $\hat{\beta}_m$ be our estimate of the magnitude of interest from the mth implicate dataset. Then the overall estimate derived using all M implicate datasets is just the average of the M separate estimates, i.e.,

$$\bar{\hat{\beta}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_m \tag{3}$$

The variance of this estimate consists of two parts. Let V_m be the variance of $\hat{\beta}_m$ estimated from the mth implicate dataset. Then the within-imputation variance *WV* is equal to the average of the *M* variances, i.e.,

$$WV = \frac{1}{M}\sum_{m=1}^{M}V_m \tag{4}$$

One would like each implicate run to explore as much as possible the domain of the joint distribution of the variables in your system; indeed, the possibility of the Markov Chain Monte Carlo process defined in (1) and (2) to jump to any part of this domain is one of the preconditions for its convergence to a joint distribution. This would imply an increased within variance, other things being equal.

The second magnitude one needs to compute is the between-imputation variance *BV*, which is given by:

$$BV = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{\beta}_m - \bar{\hat{\beta}})^2 \tag{5}$$

The between variance is an indicator of the extent to which the different implicate datasets occupy different parts of the domain of the joint distribution of the variables in our system. One would like the implicate runs to not stay far apart but rather mix with one another, thus indicating convergence to the same joint distribution. Therefore, one would like the between variance to be as small as possible relative to the within one.

The total variance *TV* of our estimate $\bar{\hat{\beta}}$ is equal to:

$$TV = WV + \frac{M+1}{M}BV \tag{6}$$

As Little and Rubin (2002) point out, the second term in (6) indicates the share of the total variance due to missing values. Having computed the total variance, one can perform a t-test of significance using the following formula to compute the degrees of freedom *df* :

$$df = (M-1)\left(1+\frac{1}{M+1}\frac{WV}{BV}\right)^2$$

Table 1.1. Distribution of journals by area and ISI code

	Research areas				
	A	**E**	**H**	**S**	**Total**
Non ISI	446	305	29	195	975
%	58.15	47.43	60.42	43.82	51.23
ISI	321	338	19	250	928
%	41.85	52.57	39.58	56.18	48.77
Total	767	643	48	445	1,903
%	100.00	100.00	100.00	100.00	100.00

Table 1.2. Distribution of journals by research area and nationality

	Research areas				
	A	**E**	**H**	**S**	**Total**
Italian	742	581	42	429	1,794
%	96.74	90.36	87.50	96.40	94.27
Other	25	62	6	16	109
%	3.26	9.64	12.50	3.60	5.73
Total	767	643	48	445	1,903
%	100.00	100.00	100.00	100.00	100.00

Table 2.1. Statistics for IF by research area

Area	mean	sd	p10	p25	p50	p75	p90	iqr
A	1.47	1.16	0.32	0.65	1.11	2.01	2.94	1.36
E	1.05	0.92	0.22	0.41	0.84	1.40	1.99	0.99
H	0.49	0.34	0.11	0.24	0.39	0.68	1.04	0.44
S	1.06	0.65	0.37	0.58	0.95	1.38	1.93	0.80
Total	1.19	0.97	0.27	0.53	0.94	1.58	2.36	1.05

Table 2.2. Statistics for IF5 by research area

Area	mean	sd	p10	p25	p50	p75	p90	iqr
A	2.44	1.90	0.76	1.17	1.94	3.02	4.92	1.85
E	1.55	1.18	0.42	0.79	1.33	2.00	2.89	1.22
H	0.73	0.36	0.34	0.44	0.63	1.12	1.24	0.67
S	1.47	0.87	0.59	0.84	1.28	1.87	2.51	1.03
Total	1.80	1.44	0.56	0.88	1.42	2.25	3.41	1.37

Table 2.3. Statistics for AIS by research area

Area	mean	sd	p10	p25	p50	p75	p90	iqr
A	0.93	1.10	0.19	0.34	0.60	0.99	2.08	0.65
E	1.09	1.54	0.17	0.35	0.64	1.06	2.34	0.71
H	0.45	0.33	0.14	0.15	0.41	0.80	0.94	0.65
S	0.95	0.69	0.31	0.51	0.72	1.23	1.89	0.72
Total	0.98	1.18	0.22	0.39	0.68	1.06	2.00	0.67

Table 2.4. Statistics for h-index by research area

Area	mean	sd	p10	p25	p50	p75	p90	iqr
A	22.77	20.78	4.00	8.00	17.00	31.00	47.00	23.00
E	21.51	18.92	4.00	7.00	16.00	30.00	47.00	23.00
H	9.31	6.34	4.00	4.00	7.00	11.50	21.00	7.50
S	19.77	16.38	4.00	7.00	14.00	28.00	43.00	21.00
Total	21.30	19.06	4.00	7.00	15.00	29.00	45.00	22.00

Table 3.1. Correlation matrix of log bibliometric indicators – Area A

	log (IF)	log(IF5)	log(AIS)	log(h)
log(IF)	1.0000			
log(IF5)	0.9192	1.0000		
log(AIS)	0.7432	0.8288	1.0000	
log(h)	0.7148	0.7636	0.7256	1.0000

Table 3.2 Correlation matrix of bibliometric indicators – Area E

	log (IF)	log(IF5)	log(AIS)	log(h)
log(IF)	1.0000			
log(IF5)	0.9592	1.0000		
log(AIS)	0.8277	0.8887	1.0000	
log(h)	0.7173	0.7753	0.7936	1.0000

Table 3.3 Correlation matrix of bibliometric indicators – Area H

	log (IF)	log(IF5)	log(AIS)	log(h)
log(IF)	1.0000			
log(IF5)	0.9323	1.0000		
log(AIS)	0.9384	0.9367	1.0000	
log(h)	0.6058	0.7164	0.6741	1.0000

Table 3.4 Correlation matrix of bibliometric indicators – Area S

	log (IF)	log(IF5)	log(AIS)	log(h)
log(IF)	1.0000			
log(IF5)	0.9272	1.0000		
log(AIS)	0.7478	0.8179	1.0000	
log(h)	0.6904	0.7290	0.6540	1.0000

Table 4.1. Prevalence of missing values for all three bibliometric indicators

Scientific areas	(1)	(2)	(3)	(4)	(5)
	Total Number of Journals	2-year Impact Factor (IF)		5-Year Impact Factor (IF5) and Article Influence Score (AIS)	
		Number of journals with a missing value	Percentage of journals with a missing value	Number of journals with a missing value	Percentage of journals with a missing value
Economics	643	319	49.61%	399	62.05%
Business, management and finance	767	447	58.28%	549	71.58%
Economic history and history of economic thought	48	30	62.50%	37	77.08%
Statistics and applied mathematics	445	195	43.82%	234	52.58%

Table 4.2. Skewness and kurtosis of the levels and logarithms of IF5 and AIS

Scientific areas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Levels				Logarithms			
	5-Year Impact Factor (IF5)		Article Influence Score (AIS)		5-Year Impact Factor (IF5)		Article Influence Score (AIS)	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
Economics	2.320	11.515	4.038	22.691	-0.674	4.179	-0.284	4.253
Business, management and finance	2.158	9.458	3.167	15.009	-0.483	4.450	-0.303	4.384
Economic history and history of economic thought	0.283	15.009	0.539	1.735	-0.351	4.384	0.054	1.433
Statistics and applied mathematics	1.526	6.500	1.938	8.397	-0.702	5.696	-1.006	7.273

Table 4.3.A. Regression Results for the 5-year Impact Factor (IF5)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline Imputation				Multiple Imputation			
	Coefficient	Std. Error	t-statistic	P value	Coefficient	Std. Error	t-statistic	P value
Panel A. Economics								
Log of h	0.952	0.069	13.750	0.000	0.019	0.148	0.130	0.899
Square of the log of h	-..-	-..-	-..-	-..-	0.012	0.023	0.520	0.606
Log of IF	-..-	-..-	-..-	-..-	0.669	0.027	24.380	0.000
Square of the log of IF	-..-	-..-	-..-	-..-	0.037	0.019	1.980	0.049
Log of AIS	-..-	-..-	-..-	-..-	0.167	0.024	6.810	0.000
Square of the log of AIS	-..-	-..-	-..-	-..-	-0.015	0.010	-1.560	0.120
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	0.086	0.061	1.420	0.158
Constant	-3.107	0.252	-12.340	0.000	0.041	0.264	0.150	0.878
Adjusted R^2		0.601				0.942		
No. of obs.		244				238		
Panel B. Business, management and finance								
Log of h	0.923	0.065	14.290	0.000	0.508	0.278	1.830	0.069
Square of the log of h	-..-	-..-	-..-	-..-	-0.049	0.039	-1.240	0.217
Log of IF	-..-	-..-	-..-	-..-	0.569	0.038	15.170	0.000
Square of the log of IF	-..-	-..-	-..-	-..-	0.082	0.031	2.670	0.008
Log of AIS	-..-	-..-	-..-	-..-	0.202	0.035	5.720	0.000
Square of the log of AIS	-..-	-..-	-..-	-..-	0.015	0.014	1.100	0.274
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	-0.103	0.246	-0.420	0.675
Constant	-2.690	0.245	-11.000	0.000	-0.570	0.579	-0.990	0.325
Adjusted R^2		0.583				0.875		
No. of obs.		218				212		
Panel C. Economic history and history of economic thought								
Log of h	1.017	0.272	3.730	0.005	-..-	-..-	-..-	-..-
Square of the log of h	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Log of IF	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Square of the log of IF	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Log of AIS	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Square of the log of AIS	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Constant	-3.144	0.823	-3.820	0.004	-..-	-..-	-..-	-..-
Adjusted R^2		0.513				-..-		
No. of obs.		11				-..-		
Panel D. Statistics and applied mathematics								
Log of h	0.778	0.061	12.780	0.000	-0.006	0.268	-0.020	0.983
Square of the log of h	-..-	-..-	-..-	-..-	0.020	0.040	0.500	0.619
Log of IF	-..-	-..-	-..-	-..-	0.661	0.038	17.330	0.000
Square of the log of IF	-..-	-..-	-..-	-..-	0.061	0.045	1.360	0.176
Log of AIS	-..-	-..-	-..-	-..-	0.225	0.032	7.080	0.000
Square of the log of AIS	-..-	-..-	-..-	-..-	0.028	0.020	1.400	0.165
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	0.102	0.096	1.060	0.292
Constant	-2.342	0.209	-11.200	0.000	-0.044	0.452	-0.100	0.923
Adjusted R^2		0.531				0.873		
No. of obs.		211				205		

Table 4.3.B. Regression Results for the Article Influence Score (AIS)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline Imputation				Multiple Imputation			
	Coefficient	Std. Error	t-statistic	P value	Coefficient	Std. Error	t-statistic	P value
Panel A. Economics								
Log of h	1.312	0.095	13.780	0.000	0.239	0.301	0.790	0.428
Square of the log of h	-..-	-..-	-..-	-..-	0.024	0.047	0.500	0.615
Log of IF	-..-	-..-	-..-	-..-	-0.312	0.134	-2.330	0.021
Square of the log of IF	-..-	-..-	-..-	-..-	-0.107	0.079	-1.350	0.177
Log of IF5	-..-	-..-	-..-	-..-	1.107	0.145	7.650	0.000
Square of the log of IF5	-..-	-..-	-..-	-..-	0.240	0.087	2.750	0.007
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	0.374	0.155	2.420	0.016
Constant	-4.983	0.339	-14.690	0.000	-2.228	0.518	-4.300	0.000
Adjusted R^2		0.630				0.803		
No. of obs.		244				238		
Panel B. Business, management and finance								
Log of h	1.135	0.104	10.910	0.000	0.315	0.513	0.610	0.540
Square of the log of h	-..-	-..-	-..-	-..-	-0.011	0.073	-0.150	0.879
Log of IF	-..-	-..-	-..-	-..-	-0.081	0.108	-0.750	0.452
Square of the log of IF	-..-	-..-	-..-	-..-	0.053	0.075	0.710	0.479
Log of IF5	-..-	-..-	-..-	-..-	0.800	0.147	5.440	0.000
Square of the log of IF5	-..-	-..-	-..-	-..-	0.093	0.072	1.280	0.201
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	1.850	0.468	3.950	0.000
Constant	-4.612	0.389	-11.850	0.000	-3.947	0.999	-3.950	0.000
Adjusted R^2		0.527				0.712		
No. of obs.		218				212		
Panel C. Economic history and history of economic thought								
Log of h	1.389	0.379	3.660	0.005	-..-	-..-	-..-	-..-
Square of the log of h	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Log of IF	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Square of the log of IF	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Log of IF5	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Square of the log of IF5	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	-..-	-..-	-..-	-..-
Constant	-4.748	1.121	-4.240	0.002	-..-	-..-	-..-	-..-
Adjusted R^2		0.455				-..-		
No. of obs.		11				-..-		
Panel D. Statistics and applied mathematics								
Log of h	0.871	0.106	8.240	0.000	0.979	0.563	1.740	0.084
Square of the log of h	-..-	-..-	-..-	-..-	-0.139	0.084	-1.640	0.102
Log of IF	-..-	-..-	-..-	-..-	-0.161	0.125	-1.290	0.199
Square of the log of IF	-..-	-..-	-..-	-..-	-0.191	0.108	-1.770	0.078
Log of IF5	-..-	-..-	-..-	-..-	0.919	0.145	6.330	0.000
Square of the log of IF5	-..-	-..-	-..-	-..-	0.155	0.120	1.290	0.199
Indicator for whether the journal content is in English	-..-	-..-	-..-	-..-	0.117	0.204	0.570	0.567
Constant	-3.174	0.356	-8.910	0.000	-2.291	0.940	-2.440	0.016
Adjusted R^2		0.428				0.617		
No. of obs.		211				205		

Table 4.4. Journal Rankings using the Baseline and the Multiple Imputation Methods

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	5-Year Impact Factor (IF5)					Article Influence Score (AIS)				
	Multiple Imputation Level 1	Multiple Imputation Level 2	Multiple Imputation Level 3	Multiple Imputation Level 4	Total	Multiple Imputation Level 1	Multiple Imputation Level 2	Multiple Imputation Level 3	Multiple Imputation Level 4	Total
Panel A. Economics										
Baseline Imputation Level 1	286	26	16	7	335	280	21	16	7	324
Baseline Imputation Level 2	22	15	12	2	51	32	22	10	4	68
Baseline Imputation Level 3	14	22	79	14	129	9	21	84	9	123
Baseline Imputation Level 4	0	1	22	105	128	1	1	18	108	128
Total	322	64	129	128	643	322	65	128	128	643
Panel B. Business, management and finance										
Baseline Imputation Level 1	350	26	9	5	390	317	38	24	10	389
Baseline Imputation Level 2	20	32	18	1	71	40	21	10	1	72
Baseline Imputation Level 3	14	17	100	22	153	26	16	89	26	157
Baseline Imputation Level 4	0	2	26	125	153	1	2	30	116	149
Total	384	77	153	153	767	384	77	153	153	767
Panel C. Statistics and applied mathematics										
Baseline Imputation Level 1	209	15	7	3	234	196	15	15	6	232
Baseline Imputation Level 2	11	18	6	1	36	18	10	7	0	35
Baseline Imputation Level 3	3	11	70	2	86	9	19	55	6	89
Baseline Imputation Level 4	0	0	6	83	89	0	0	12	77	89
Total	223	44	89	89	445	223	44	89	89	445

Notes**: Higher level numbers denote higher-ranked journals**.

Table 4.5. Differences in Journal Rankings between the Baseline and the Multiple Imputation Methods

Difference in Ranking across Imputation Methods	(1)	(2)	(3)	(4)
	5-Year Impact Factor (IF5)		Article Influence Score (AIS)	
	Number of Journals	Percentage of All Journals	Number of Journals	Percentage of All Journals
Panel A. Economics				
Ranking difference = - 3	7	1.09%	7	1.09%
Ranking difference = - 2	18	2.80%	20	3.11%
Ranking difference = - 1	52	8.09%	40	6.22%
Ranking difference = 0	485	75.43%	494	76.83%
Ranking difference = + 1	66	10.26%	71	11.04%
Ranking difference = + 2	15	2.33%	10	1.56%
Ranking difference = + 3	0	0.00%	1	0.16%
Percentage of journals for which the difference in ranking is between -1 and 1	93.78%		94.09%	
Panel B. Business, management and finance				
Ranking difference = - 3	5	0.65%	10	1.30%
Ranking difference = - 2	10	1.30%	25	3.26%
Ranking difference = - 1	66	8.61%	74	9.65%
Ranking difference = 0	607	79.14%	543	70.80%
Ranking difference = + 1	63	8.21%	86	11.21%
Ranking difference = + 2	16	2.09%	28	3.65%
Ranking difference = + 3	0	0.00%	1	0.13%
Percentage of journals for which the difference in ranking is between -1 and 1	95.96%		91.66%	
Panel C. Statistics and applied mathematics				
Ranking difference = - 3	3	0.67%	6	1.35%
Ranking difference = - 2	8	1.80%	15	3.37%
Ranking difference = - 1	23	5.17%	28	6.29%
Ranking difference = 0	380	85.39%	338	75.96%
Ranking difference = + 1	28	6.29%	49	11.01%
Ranking difference = + 2	3	0.67%	9	2.02%
Ranking difference = + 3	0	0.00%	0	0.00%
Percentage of journals for which the difference in ranking is between -1 and 1	96.85%		93.26%	

Table 4.6. Journal Rankings across Bibliometric Indicators, Baseline Imputation Method

IF5 Level	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	AIS Level	(11)	(12)	(13)	(14)	(15)
	IF5 versus AIS					IF5 versus h-index						AIS versus h-index				
	AIS Level 1	AIS Level 2	AIS Level 3	AIS Level 4	Total	h-index Level 1	h-index Level 2	h-index Level 3	h-index Level 4	Total		h-index Level 1	h-index Level 2	h-index Level 3	h-index Level 4	Total
Panel A. Economics																
IF5 Level 1	321	11	3	0	335	313	11	11	0	335	**AIS Level 1**	313	5	6	0	324
IF5 Level 2	1	40	9	1	51	14	24	11	2	51	**AIS Level 2**	18	41	6	3	68
IF5 Level 3	1	16	89	23	129	6	27	75	21	129	**AIS Level 3**	3	16	88	16	123
IF5 Level 4	1	1	22	104	128	1	1	30	96	128	**AIS Level 4**	0	1	27	100	128
Total	324	68	123	128	643	334	63	127	119	643	**Total**	334	63	127	119	643
Panel B. Business, management and finance																
IF5 Level 1	381	6	2	1	390	375	6	9	0	390	**AIS Level 1**	376	6	7	0	389
IF5 Level 2	6	58	4	3	71	14	42	11	4	71	**AIS Level 2**	12	44	12	4	72
IF5 Level 3	0	6	132	15	153	2	23	114	14	153	**AIS Level 3**	2	22	110	23	157
IF5 Level 4	2	2	19	130	153	2	3	17	131	153	**AIS Level 4**	3	2	22	122	149
Total	389	72	157	149	767	393	74	151	149	767	**Total**	393	74	151	149	767
Panel C. Economic history and history of economic thought																
IF5 Level 1	25	1	0	0	26	25	0	1	0	26	**AIS Level 1**	25	0	0	0	25
IF5 Level 2	0	3	0	0	3	0	3	0	0	3	**AIS Level 2**	0	3	2	0	5
IF5 Level 3	0	1	8	1	10	0	2	6	2	10	**AIS Level 3**	0	2	4	3	9
IF5 Level 4	0	0	1	8	9	0	0	2	7	9	**AIS Level 4**	0	0	3	6	9
Total	25	5	9	9	48	25	5	9	9	48	**Total**	25	5	9	9	48
Panel D. Statistics and applied mathematics																
IF5 Level 1	219	7	7	1	234	213	11	10	0	234	**AIS Level 1**	214	9	7	2	232
IF5 Level 2	6	20	10	0	36	6	21	7	2	36	**AIS Level 2**	3	24	6	2	35
IF5 Level 3	7	6	51	22	86	3	17	44	22	86	**AIS Level 3**	4	13	41	31	89
IF5 Level 4	0	2	21	66	89	1	3	21	64	89	**AIS Level 4**	2	6	28	53	89
Total	232	35	89	89	445	223	52	82	88	445	**Total**	223	52	82	88	445

Table 4.7. Differences in Journal Rankings across Bibliometric Indicators, Baseline Imputation Method

Difference in Ranking across Imputation Methods	(1)	(2)	(3)	(4)	(5)	(6)
	IF5 versus AIS		IF5 versus h-index		AIS versus h-index	
	Number of Journals	Percentage of All Journals	Number of Journals	Percentage of All Journals	Number of Journals	Percentage of All Journals
Panel A. Economics						
Ranking difference = - 3	0	0.00%	0	0.00%	0	0.00%
Ranking difference = - 2	4	0.62%	13	2.02%	9	1.40%
Ranking difference = - 1	43	6.69%	43	6.69%	27	4.20%
Ranking difference = 0	554	86.16%	508	79.01%	542	84.29%
Ranking difference = + 1	39	6.07%	71	11.04%	61	9.49%
Ranking difference = + 2	2	0.31%	7	1.09%	4	0.62%
Ranking difference = + 3	1	0.16%	1	0.16%	0	0.00%
Percentage of journals for which the difference in ranking is between -1 and 1	98.91%		96.73%		97.98%	
Panel B. Business, management and finance						
Ranking difference = - 3	1	0.13%	0	0.00%	0	0.00%
Ranking difference = - 2	5	0.65%	13	1.70%	11	1.43%
Ranking difference = - 1	25	3.26%	31	4.04%	41	5.35%
Ranking difference = 0	701	91.40%	662	86.31%	652	85.01%
Ranking difference = + 1	31	4.04%	54	7.04%	56	7.30%
Ranking difference = + 2	2	0.26%	5	0.65%	4	0.52%
Ranking difference = + 3	2	0.26%	2	0.26%	3	0.39%
Percentage of journals for which the difference in ranking is between -1 and 1	98.70%		97.39%		97.65%	
Panel C. Economic history and history of economic thought						
Ranking difference = - 3	0	0.00%	0	0.00%	0	0.00%
Ranking difference = - 2	0	0.00%	1	2.08%	0	0.00%
Ranking difference = - 1	2	4.17%	2	4.17%	5	10.42%
Ranking difference = 0	44	91.67%	41	85.42%	38	79.17%
Ranking difference = + 1	2	4.17%	4	8.33%	5	10.42%
Ranking difference = + 2	0	0.00%	0	0.00%	0	0.00%
Ranking difference = + 3	0	0.00%	0	0.00%	0	0.00%
Percentage of journals for which the difference in ranking is between -1 and 1	100.00%		97.92%		100.00%	
Panel D. Statistics and applied mathematics						
Ranking difference = - 3	1	0.23%	0	0.00%	2	0.45%
Ranking difference = - 2	7	1.57%	12	2.70%	9	2.02%
Ranking difference = - 1	39	8.76%	40	8.99%	46	10.34%
Ranking difference = 0	356	80.00%	342	76.85%	332	74.61%
Ranking difference = + 1	33	7.42%	44	9.89%	44	9.89%
Ranking difference = + 2	9	2.02%	6	1.35%	10	2.25%
Ranking difference = + 3	0	0.00%	1	0.23%	2	0.45%
Percentage of journals for which the difference in ranking is between -1 and 1	96.18%		95.73%		94.83%	

Table 6.1

Ranking list of Italian journals

	Area	h	cl_ais	cl_if5	cl_fin
contabilità e cultura aziendale	A	3	D	D	D
sanità pubblica e privata	A	3	D	D	D
analisi giuridica dell'economia	A	3	D	D	D
economia dei servizi	A	3	D	D	D
economia e diritto del terziario	A	3	D	D	D
impresa progetto - electronic journal of management	A	3	D	D	D
bancaria	A	3	D	D	D
finanza marketing e produzione	A	4	D	D	D
mecosan	A	4	D	D	D
esperienze d'impresa	A	4	D	D	D
mercati e competitività	A	4	D	D	D
sviluppo & organizzazione	A	4	D	D	D
studi organizzativi	A	4	D	D	D
piccola impresa	A	4	D	D	D
riv. rassegna italiana di valutazione	A	5	D	D	D
sinergie	A	5	D	D	D
economia della cultura	A	5	D	D	D
rivista dei dottori commercialisti	A	5	D	D	D
rivista italiana di ragioneria e di economia aziendale	A	5	D	D	D
economia & management	A	6	D	D	C
banca impresa società	A	6	D	D	C
symphonya emerging issues in management	A	6	D	D	C
micro & macro marketing	A	6	D	D	C
mercato concorrenza regole	A	6	D	D	C
azienda pubblica	A	7	D	D	C
rivista italiana di economia, demografia e statistica	E	3	D	D	D
rassegna economica	E	3	D	D	D
il ponte	E	3	D	D	D
rivista della cooperazione	E	3	D	D	D
rivista trimestrale di diritto dell“economia	E	3	D	D	D
economia marche	E	3	D	D	D
studi sull'integrazione europea	E	3	D	D	D
federalismo fiscale	E	3	D	D	D
argomenti	E	3	D	D	D
studi economici	E	3	D	D	D
european transport\trasporti europei	E	3	D	D	D
economia & politica del farmaco	E	3	D	D	D
quaderni di rassegna sindacale	E	3	D	D	D
istituzioni e sviluppo economico	E	4	D	D	D
politica agricola internazionale	E	4	D	D	D
prospettive sociali e sanitarie	E	4	D	D	D
il diritto dell'economia	E	4	D	D	D
economia & diritto agroalimentare	E	4	D	D	D
politiche sanitarie	E	4	D	D	D
rivista di economia e statistica del territorio	E	4	D	D	D
economia delle fonti di energia e dell'ambiente	E	4	D	D	D
sviluppo locale	E	4	D	D	D
economia & lavoro	E	4	D	D	D

economia agro-alimentare	E	4	D	D	D
il fisco	E	4	D	D	D
qa – rivista dell'associazione rossi-doria	E	5	D	D	D
economia e società regionale	E	5	D	D	D
diritto immigrazione e cittadinanza	E	5	D	D	D
rivista di diritto finanziario e scienza delle finanze	E	5	D	D	D
diritto ed economia	E	5	D	D	D
sociologia e politiche sociali	E	5	D	D	D
moneta e credito	E	5	D	D	D
rivista internazionale di scienze ec. e comm.	E	5	D	D	D
studi e note di economia	E	5	D	D	D
rivista internazionale di scienze sociali	E	5	D	D	D
agricoltura istituzioni mercati	E	5	D	D	D
economia pubblica	E	5	D	D	D
archivio di studi urbani e regionali	E	5	D	D	D
rivista di diritto ed economia dello sport	E	5	D	D	D
new medit	E	5	D	D	D
economia italiana	E	6	D	D	D
economia internazionale	E	6	D	D	D
impresa sociale	E	6	D	D	D
scienze regionali	E	7	D	D	D
rivista di economia agraria	E	7	D	D	D
psl quarterly review (formerly bnl quarterly review)	E	8	D	D	D
rivista economica del mezzogiorno	E	8	D	D	D
economia e politica industriale	E	9	D	D	D
lavoro e diritto	E	9	D	D	D
rivista italiana di politiche pubbliche	E	9	D	D	D
la rivista delle politiche sociali	E	9	D	D	D
l'industria	E	10	D	D	C
politica economica	E	10	D	D	C
polis	E	10	D	D	C
rivista italiana degli economisti	E	10	D	D	C
economia politica	E	10	D	D	C
giornale degli economisti e annali di economia	E	11	D	D	C
stato e mercato	E	12	D	D	C
rivista di politica economica	E	12	D	D	C
research in economics	E	13	D	D	C
economic notes	E	16	C	C	C
labour	E	26	B	B	B
storia del pensiero economico	H	4	D	D	D
popolazione e storia	H	4	D	D	D
il pensiero economico italiano	H	4	D	D	D
journal of european economic history	H	5	D	D	D
history of economic ideas	H	7	D	D	C
rivista di storia economica	H	8	C	C	C
rendiconti per gli studi economici quantitativi	S	3	D	D	D
rivista di statistica ufficiale	S	3	D	D	D
rivista di studi familiari	S	3	D	D	D
giornale dell'istituto italiano degli attuari	S	4	D	D	D
tpm. testing psicometria metodologia	S	4	D	D	D
informatica economica	S	4	D	D	D
statistica	S	4	D	D	D
studi emigrazione	S	4	D	D	D
statistica & applicazioni	S	5	D	D	D
igiene e sanità pubblica	S	5	D	D	D

quaderni di statistica	S	5	D	D	D
statistica applicata	S	5	D	D	D
decisions in economics and finance	S	9	D	D	C
ricerche di matematica	S	9	D	D	C
metron	S	9	D	D	C
epidemiologia e prevenzione	S	14	D	D	C

Table 7.1

Final classification of Italian Journals

	Area				
cl_fin	**A**	**E**	**H**	**S**	**Total**
Level 1 (D)	19	51	4	12	86
%	76.00	82.26	66.67	75.00	78.90
Level 2 (C)	6	10	2	4	22
%	24.00	16.13	33.33	25.00	20.18
Level 3 (B)	0	1	0	0	1
	0.00	1.61	0.00	0.00	0.92
Total	25	62	6	16	109
	100.00	100.00	100.00	100.00	100.00

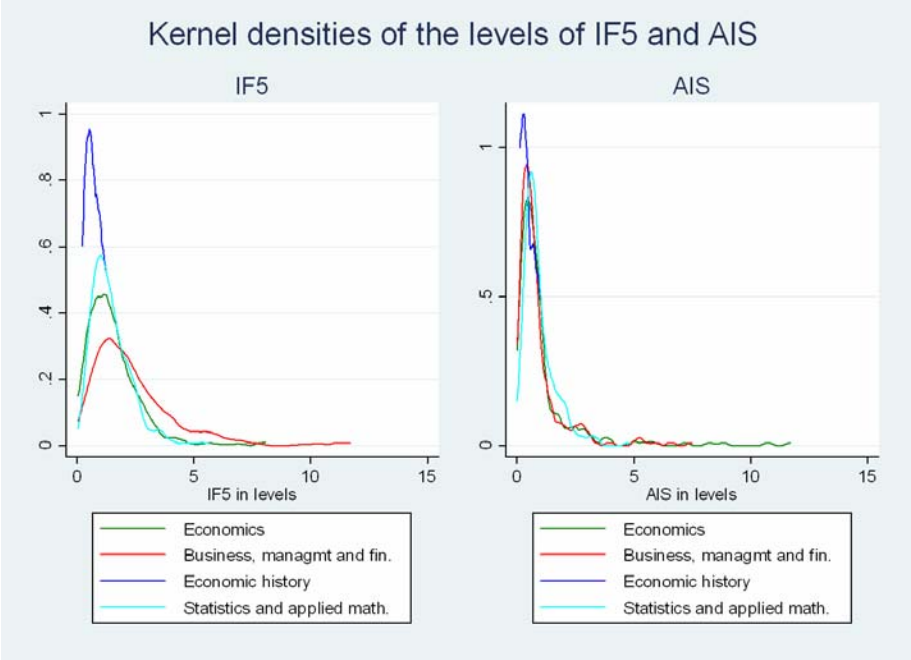
Table 7.2

Final classification of journals

	Area				
	A	**E**	**H**	**S**	**Total**
Level 1 (D)	375	312	24	215	926
%	48.89	48.52	50.00	48.31	48.66
Level 2 (C)	76	61	5	37	179
%	9.91	9.49	10.42	8.31	9.41
Level 3 (B)	144	118	9	81	352
%	18.77	18.35	18.75	18.20	18.50
Level 4 (A)	172	152	10	112	446
%	22.43	23.64	20.83	25.17	23.44
Total	767	643	48	445	1,903
	100.00	100.00	100.00	100.00	100.00

Figure 4.1. Kernel densities of the levels and logarithms of IF5 and AIS, non-missing values

A. Levels



B. Logarithms

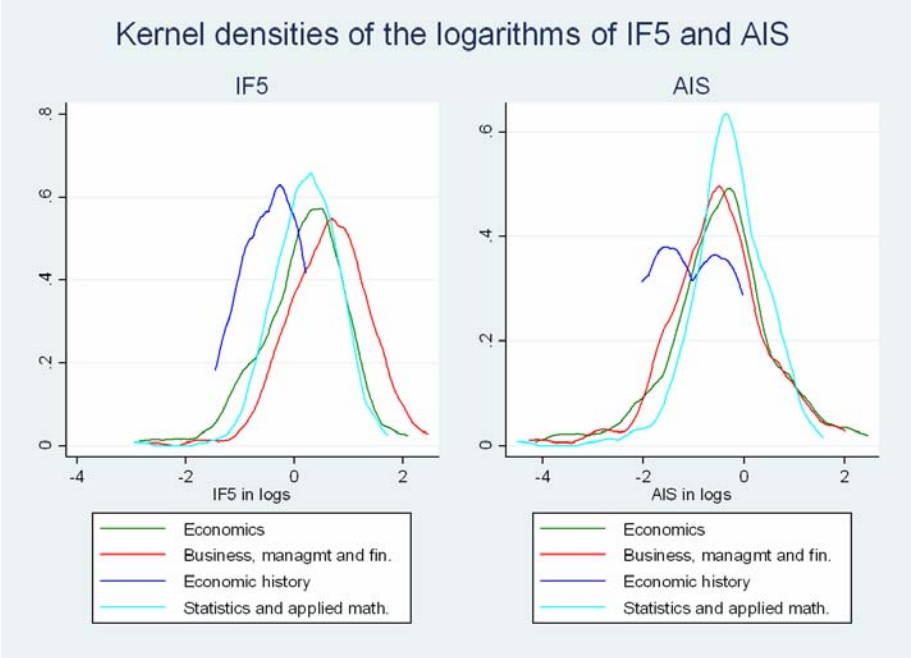


Figure 4.3. Imputed versus non-imputed values of the logarithm of AIS in economics, selected implicates

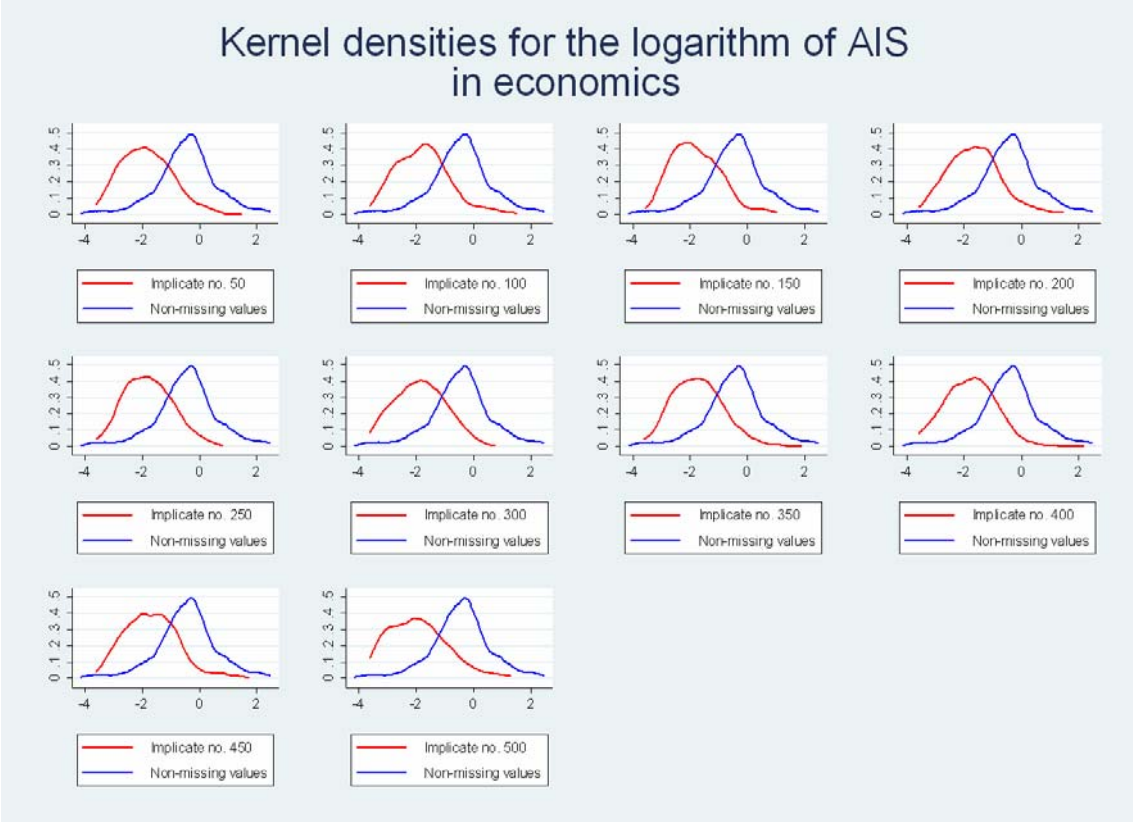
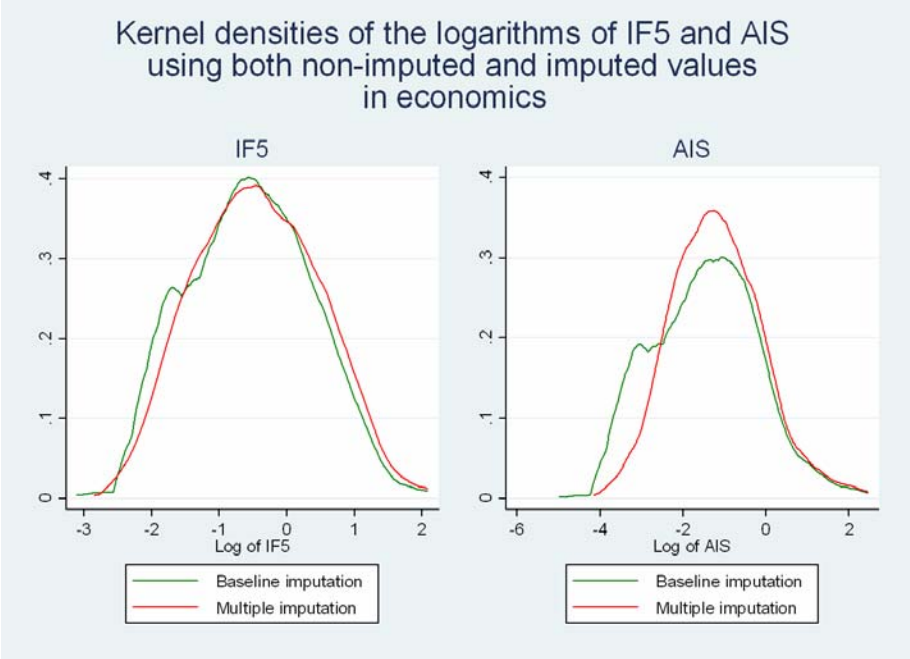


Figure 4.4. Non-imputed and imputed values from Baseline and Multiple Imputation Methods, logarithms

A. Economics



B. Business, management and finance

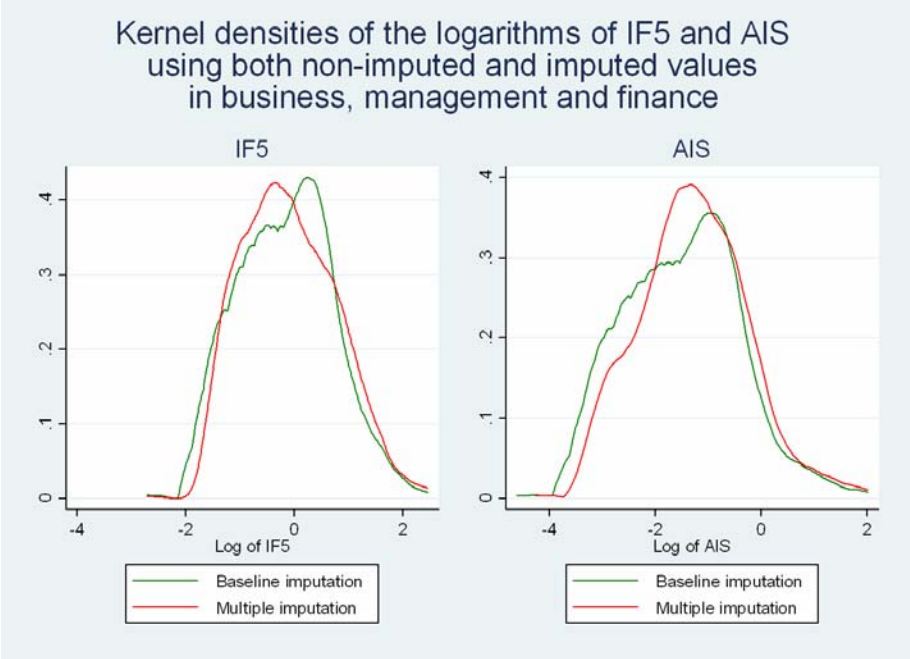
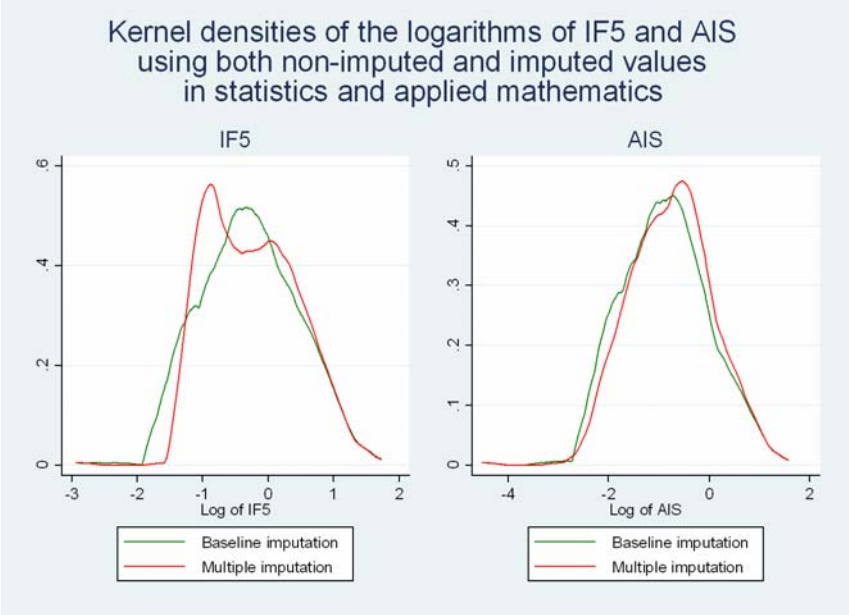


Figure 4.4. Non-imputed and Imputed Values from Baseline and Multiple Imputation Methods, Logarithms (continued)

C. Statistics and applied mathematics



Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Appendice D

Analisi delle citazioni

Le citazioni sono disponibili per 3.475 articoli pubblicati nel 2004-2010 in riviste censite in WoS. Tra questi, 1.530 sono pubblicati su riviste dell'area economica, 1.178 di area statistica, 681 di area aziendale, 75 di storia economica e 11 su riviste generaliste (Tab. 1). La Tab. 1 indica anche la forte concentrazione delle citazioni, che raggiungono valori elevati solo per il primo decile della distribuzione.

La distribuzione delle citazioni per le 4 sotto aree di classificazione (Economia -E-, Economia Aziendale -A-, Statistica -S- e Storia -H-) rivela una leggera differenza tra le aree, con gli articoli di Economia Aziendale che hanno la media di citazioni più alta (4,22), escludendo gli articoli pubblicati su riviste generaliste (G), e gli articoli di Storia Economica che hanno la media di citazioni più bassa (1,09).

Tab. 1: Distribuzione delle citazioni nel 2004-2010 degli articoli presentati per la VQR pubblicati in riviste censite WoS.

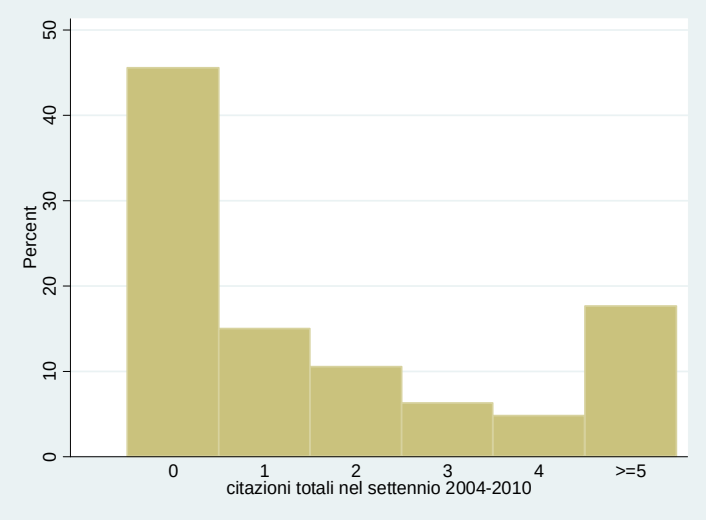
area	*media*	*sd*	*p10*	*p25*	*p50*	*p75*	*p90*	*max*	*N*
A	4,22	9,04	0,00	0,00	1,00	4,00	11,00	72,00	681
E	3,04	7,77	0,00	0,00	1,00	3,00	8,00	111,00	1.530
G	9,45	9,51	0,00	1,00	9,00	12,00	12,00	34,00	11
H	1,09	2,67	0,00	0,00	0,00	1,00	2,00	15,00	75
S	2,68	5,87	0,00	0,00	1,00	3,00	7,00	99,00	1.178
Totale	3,13	7,43	0,00	0,00	1,00	3,00	8,00	111,00	3.475

La distribuzione delle citazioni indica che circa metà degli articoli pubblicati nel 2004-2010 sulle riviste censite da WoS non ha alcuna citazione nello stesso periodo; solo 614 articoli (il 18%) ha 5 o più citazioni (Tab. 2 e Fig.1).

Tab. 2: Articoli per numero di citazioni nel 2004-2010

citazioni totali	*# articoli*	*%*
0	1.584	45,58
1	523	15,05
2	367	10,56
3	219	6,30
4	168	4,83
>5	614	17,67
Totale	3.475	100,00

Fig. 1: Distribuzione delle citazioni nel 2004-2010



La Tab. 3 indica che vi è una forte correlazione tra ranking delle riviste e numero di citazioni. Dei 614 articoli con almeno 5 citazioni, 513 sono in classe Eccellente (84%), 77 in classe Buono, 17 in classe Accettabile e 7 in classe Limitato.

Tab 3: Distribuzione nelle 4 classi di merito degli articoli con almeno 5 citazioni nel 2004-2010

classe bibliometrica	*# articoli*	*%*
Eccellente	513	83,55
Buono	77	12,54
Accettabile	17	2,77
Limitato	7	1,14
Totale	614	100,00

Come previsto dai criteri di valutazione, la classe di merito finale di ciascun articolo dipende dalla classe di merito iniziale della rivista (calcolata sulla base di indicatori bibliometrici) e dal numero di citazioni ricevute da ciascun articolo in WoS, solo per gli articoli con un "numero significativo di citazioni". La soglia che identifica gli articoli con un "numero significativo di citazioni" è stata determinata in 5 citazioni annue. Tale valore è pari (approssimativamente) al numero di citazioni ricevute in media nel 2011 da lavori pubblicati nel 2005-06 su riviste delle subject list WoS rilevanti per l'area 13.

In pratica, se un articolo ha ricevuto un "numero significativo di citazioni nelle riviste WoS nel 2004-2010 in rapporto agli anni trascorsi dalla pubblicazione", è stato "promosso" di una classe. A causa della mancanza di dati affidabili per l'analisi delle citazioni dei singoli articoli, nessuna correzione è stata applicata per gli articoli pubblicati in riviste "non WoS". Nessuna "retrocessione" è stata applicata alla classe iniziale di merito di un prodotto di ricerca.

Gli articoli con oltre 4 citazioni medie annue sono 93 (Tab. 4) di cui 87 già classificati in classe Eccellente dagli indicatori bibliometrici delle riviste; di conseguenza, in seguito all'analisi delle citazioni, 6 articoli sono stati promossi dalla classe Buono alla classe Eccellente (Tab.5).

Tab 4: Numero di articoli per classi di citazioni medie annue

citazioni medie annue (c)	*Economia*	*Economia Aziendale*	*Statistica*	*Storia*	*Generaliste*	*Totale*
0	730	275	533	47	2	1.587
0≤c<1	542	237	460	24	1	1.264
1≤c<2	141	80	117	1	0	339
2≤c<3	51	41	31	3	3	129
3≤c<4	28	14	20	0	1	63
c>4	38	34	17	0	4	93
Totale	1.530	681	1.178	75	11	3.475

Tab 5: Numero di articoli con oltre 4 citazioni medie annue

bibliometria senza citazioni	***# articoli***	***%***
Eccellente	87	93,55
Buono	6	6,45
Total	93	100,00

Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Appendice E

La valutazione peer review e i revisori esterni

1. La valutazione peer review

Il GEV13 ha valutato in peer review 6.277 lavori (52,6% dei lavori complessivamente valutati). Essi comprendono:

- tutti i prodotti di ricerca diversi da articoli su riviste (monografie, capitoli di libro, atti di congresso, ecc.);
- tutti gli articoli su riviste non classificate dal GEV13 nell'elenco pubblicato sul sito Anvur il 4 settembre 2012;
- un campione casuale stratificato per area di ricerca (Economia -E Storia -H-, Economia Aziendale -M-, , Statistica -S-) degli articoli su riviste classificate dal GEV13 (590 articoli su rivista).

Ciascun prodotto da valutare in peer review è stato assegnato ad uno dei 3 sub-GEV in base al SSD e poi ripartito tra i membri del GEV in base all'argomento principale della ricerca. All'interno di ciascun sub-GEV (Economia, Economia Aziendale e Finanza, Statistica e Matematica Applicata), sono state definite 33 aree di ricerca - *subject categories* in Tab. 1 - per facilitare la classificazione e la successiva assegnazione dei prodotti ai membri GEV e ai revisori.

La gestione della peer review è avvenuta in modo del tutto simile a quella di una rivista, con la differenza importante che ciascun prodotto è stato preso in carico da due membri GEV, e non da un solo editor come avviene normalmente per le riviste scientifiche.

Ciascun prodotto da valutare in peer review è stato dunque assegnato dai coordinatori dei sub-GEV a due membri dello stesso sub-GEV che non sono stati rivelati l'uno all'altro (ad eccezione di quanto notato più avanti per quanto riguarda i prodotti dei membri del GEV). In altre parole, i due membri GEV responsabili del prodotto non conoscevano l'identità reciproca fino al termine della valutazione. Inoltre, ciascuna subject category è stata affidata ad almeno quattro diversi membri GEV (talvolta a più di quattro); anche in questo caso l'identità dei membri GEV non era nota. L'assegnazione dei prodotti ai membri GEV ha evitato tutti i conflitti di interesse con gli autori dei prodotti e con la sede di ricerca degli autori. Sono stati inoltre esclusi casi di potenziali conflitti di interesse per progetti di ricerca comune, e affiliazioni precedenti dei membri GEV con gli autori e degli autori, in linea con i criteri indicati dal GEV.

I prodotti di ricerca dei membri GEV e del Presidente GEV sono stati valutati seguendo le regole previste dalla VQR (Linee Generali per i Gruppi di Esperti della VQR, Sezione 5). In particolare, il Presidente ha assegnato a revisori esterni i prodotti di ricerca dei componenti del GEV (in caso di prodotti di tipologia diversa da articoli su riviste classificate dal GEV o prodotti del campione casuale degli articoli su riviste classificate) e l'ANVUR ha gestito i prodotti di ricerca del Presidente del GEV. Va evidenziato che questi sono stati gli unici casi di prodotti assegnati ad uno solo membro GEV (il Presidente o il Coordinatore della VQR).

Tab. 1 : Subject category definite dal GEV per la distribuzione dei lavori ai membri GEV e ai revisori.

sub-GEV	Subject categories
Economia	1. Microeconomics 2. Macroeconomics and monetary economics 3. International economics 4. Public economics 5. Health, education and welfare 6. Labor economics 7. Industrial organization 8. Economic development, technical change and growth 9. Environmental and ecological economics 10. Urban, rural, regional and transportation economics 11. History of economic thought 12. Economic history
Economia Aziendale e Finanza	13. Business administration 14. Business economics 15. Financial accounting 16. Managerial accounting 17. Public accounting and management 18. Marketing 19. Corporate finance 20. Financial markets and asset pricing 21. Organization 22. Strategy 23. Operations 24. Technology and Innovation Management 25. Commodity science
Statistica e Matematica Applicata	26. Econometrics; Economic and business statistics 27. Data collection and production; Survey methodology 28. Statistical modeling and inference; Statistical decision theory 29. Classification and data analysis; Multivariate statistics 30. Biostatistics 31. Demography; Social statistics 32. Mathematical economics 33. Financial and insurance mathematics

In media, ciascun membro GEV è stato responsabile di 338 prodotti; il numero è maggiore nell'area aziendale (432 lavori), che ha registrato il numero maggiore di lavori in peer review. I membri GEV hanno effettuato una prima valutazione interna dei prodotti per individuare i "prodotti non valutabili" (prodotti appartenenti a tipologie escluse dalla VQR, prodotti con allegati mancanti o non adeguati per la valutazione, prodotti pubblicati in anni precedenti o successivi al periodo di riferimento della VQR, lavori non pubblicati, lavori senza alcun contenuto di ricerca come libri di esercizi per il triennio, working papers) e i prodotti di qualità chiaramente limitata.

I prodotti sono stati valutati internamente se entrambi i membri GEV – con scelta del tutto indipendente – avevano valutato i prodotti come "nettamente al di sotto del 50% della scala di valore condivisa dalla comunità internazionale". Complessivamente, il GEV ha valutato internamente con questa modalità (quindi in classe D) 653 lavori (il 10.8% dei lavori complessivamente valutati in peer review) per cui vi era consenso unanime.

In tutti gli altri casi i prodotti sono stati inviati alla valutazione esterna. In particolare, i due membri GEV che avevano in carico ciascun prodotto hanno scelto ciascuno un revisore esterno al GEV in maniera indipendente. La scelta dei revisori è stata fatta evitando conflitti di interesse tra revisori ed autori e tra revisori e sede di ricerca degli autori, anche in questo caso seguendo le regole della VQR. L'indipendenza dei revisori è stata garantita prestando attenzione alla sede di ricerca dei revisori, alle collaborazioni di ricerca e, ove possibile, alla nazionalità. Nei casi in cui il revisore abbia evidenziato la presenza di conflitti di interessi non noti al GEV, la non competenza sullo specifico prodotto, e la mancata disponibilità per motivi di tempo, il membro GEV che aveva proceduto all'assegnazione ha provveduto a modificarla e assegnare nuovamente il prodotto ad un altro revisore.

I revisori hanno fornito una valutazione qualitativa dei prodotti di ricerca utilizzando una scheda di valutazione predisposta dal GEV (Tab. 2) e costituita da una serie di tre domande tendenti a valutare le caratteristiche di rilevanza, originalità, internazionalizzazione previste dal bando e da un campo libero con numero limitato di parole (500). Il punteggio di sintesi è stato ottenuto sommando i punteggi delle tre rispose, e attribuendo una classe di merito complessiva, come indicato in Tab. 3. Dato l'elevato numero di revisori stranieri, la scheda di valutazione è stata predisposta in inglese per tutti i prodotti valutati.

Tab.2: Scheda per la valutazione qualitativa dei prodotti di ricerca in peer review.

ANVUR – ASSESSMENT OF THE RESEARCH QUALITY 2004-2010

Assessment Form (one form to be filled for each research product)

Groups of Experts for Economics and Statistics - GEV 13.

In the following research output or work means: journal article, book chapter, monograph, conference proceeding. For each of the 3 criteria (relevance, originality / innovativeness, international reach / impact) a non exhaustive list of questions is provided to clarify its meaning.

Q1. Relevance. Are the research questions addressed by the work of general, narrow or limited interest? Are they likely to spur additional work? Are the methods, the data or the results likely to be used by other researchers?

Please grade the research output in terms of its relevance, expressing a score between 1 and 9, with **1 and 9 indicating minimal and maximal relevance**, respectively.

1 2 3 4 5 6 7 8 9

Q2. Originality / innovativeness. Does the work advance knowledge in some dimension? Does it pose new questions, provide new answers, use new data or methods?

Please grade the research output in terms of its originality, expressing a score between 1 and 9, with **1 and 9 indicating minimal and maximal originality / innovativeness**, respectively.

1 2 3 4 5 6 7 8 9

Q3. International reach / Impact: Was the work able to reach an international audience, or does it have the potential to do so? Was it cited, quoted or reviewed by other researchers, or do you expect it will be in the future? Is it likely to leave a mark in the international scientific community? Did the work consider the relevant international contributions on the same or related issues?

Please grade the research output in terms of its international reach and impact, expressing a score between **1 and 9, with 1 and 9 indicating minimal and maximal international reach/impact**, respectively.

1 2 3 4 5 6 7 8 9

Q4. **Optional (max. 1000 char.)** Free format explanations of the grades:
Relevance:
Originality/Innovativeness:
International reach / Impact:

Tab. 3: Corrispondenza punteggi e classi di merito nella scheda di valutazione peer review

Punteggio	Classe di merito
23-27	Eccellente / Excellent
18-22	Buono / Good
15-17	Accettabile / Acceptable
1-14	Limitato / Limited

Tab. 4: Gruppi di consenso che hanno raggiunto la valutazione di sintesi a maggioranza per sub-GEV e SSD. Non vi sono stati casi di dissenso nel sub-GEV di statistica.

Sub-GEV	SSD	# dissensi
Economia	SECS/P01	5
	SECS/P02	0
	SECS/P03	0
	SECS/P04	0
	SECS/P06	4
	SECS/P12	1
	Totale	**10**
Economia Aziendale e Finanza	SECS/P07	9
	SECS/P08	6
	SECS/P09	1
	SECS/P10	11
	SECS/P11	1
	SECS/P13	6
	Totale	**34**

Una volta ricevute le schede di valutazione dei valutatori, il GEV ha proceduto alle valutazioni di sintesi costituendo, per ciascun prodotto valutato, 6.277 Gruppi di Consenso, costituiti dai due componenti del GEV che avevano assegnato il prodotto ai revisori esterni. I Gruppi di Consenso hanno definito la valutazione di sintesi all'unanimità oppure ricorrendo (in casi limitati) ad un terzo referee e sono stati integrati, solo in caso di conflitto di valutazioni tra i componenti del Gruppo di Consenso, dai coordinatori sub-GEV o dal Presidente del GEV. Particolare attenzione da parte dei Gruppi di Consenso e dei coordinatori dei sub-GEV è stata posta all'esame di prodotti valutati da valutatori con pareri discordi, sia nel caso in cui i pareri erano fortemente in disaccordo, sia nel caso in cui erano entrambi vicino alla soglia di separazione tra le classi di merito. Come evidenziato in Tab. 3, i casi di conflitto di valutazione di sintesi sono stati complessivamente 34 su più di 6.000 prodotti valutati.

Il ricorso al terzo revisore esterno è stato limitato a soli 3 casi, grazie all'ampio consenso sulle valutazioni di sintesi nei Gruppi.

La definizione dei lavori "non valutabili" è stata particolarmente complessa ed ha richiesto molto tempo, per tre motivi. In primo luogo, il bando VQR prevede linee guida di carattere generale circa la tipologia e l'ammissibilità dei prodotti ed è stato interpretato in modo diverso dai singoli atenei e ricercatori. In secondo luogo, in alcuni casi (marginali) capire se la tipologia di pubblicazione è una rivista, una monografia o altro ancora non è un compito semplice, ed è soggetto a margini di errore. In terzo luogo, non conoscendo ex-ante la tipologia di lavori presentati alla VQR, i Criteri stessi di valutazione del GEV 13 non avevano delineato in maniera esauriente e sufficientemente chiara i casi in cui un lavoro sarebbe stato considerato "non valutabile". Questo potrebbe avere generato incertezze negli atenei a tra i ricercatori circa le tipologie "ammissibili".

I lavori individuati come "non valutabili" sono stati prima di tutto evidenziati da decisioni unanimi dei Gruppi di Consenso; successivamente sono stati esaminati dai Coordinatori sub-GEV e dal Presidente. Le motivazioni di non valutabilità sono state inviate alle strutture per consentire le opportune verifiche e controdeduzioni. A seguito delle controdeduzioni, il GEV ha considerato non valutabili 51 lavori.

Il GEV13 ha cercato di cercare di ridurre al minimo i casi di ambiguità (rientrano in questa tipologia, ad esempio, i lavori inviati più di una volta da una stessa struttura, quelli pubblicati in date diverse dal 2004-2010, ecc); nel caso ambiguità fossero ancora presenti, sono state risolte nel senso di ammettere i lavori alla valutazione. Le motivazioni di non valutabilità sono riportate in Tabella 5.

I Coordinatori sub-GEV e il Presidente GEV sono stati coinvolti anche in tutti i Gruppi di Consenso costituiti per verificare l'insieme dei "prodotti non valutabili". Le motivazioni di non ammissibilità sono state inviate alle Strutture (Università ed Enti di Ricerca) per consentire loro di le opportune verifiche e controdeduzioni.

Tab 5. Lavori non valutabili e principali motivazioni.

Motivazione di non valutabilità	# lavori
Manca pdf	11
Curatela	3
Duplicati	6
Libro di esercizi o applicazioni	11
Manca ISBN	9
Pubblicato in anni diversi dalla VQR 2004-2010	2
Traduzione	7
Mimeo	2
Totale	51

2. I revisori esterni

L'elenco dei revisori è stato costruito stabilendo standard di qualità scientifica, impatto sulla comunità scientifica internazionale, esperienza nella valutazione, competenza nelle rispettive aree di valutazione, copertura delle aree di ricerca dell'area 13. Per limitare conflitti di interesse, il GEV ha cercato di coinvolgere, per quanto possibile, revisori di istituzioni straniere e di enti di ricerca non sottoposti alla valutazione.

Dal punto di vista operativo, il Presidente ed i coordinatori sub-GEV hanno inizialmente consultato la lista dei revisori di Area 13 del CINECA e l'elenco dei revisori che si sono autonomamente proposti attraverso il sito del CINECA; hanno inoltre chiesto ai membri GEV di proporre, indipendentemente gli uni dagli altri, un numero minimo di nominativi (15-20) per le subject categories di competenza. Ogni nominativo proposto è stato accompagnato da informazioni sulla ricerca scientifica del potenziale revisore fornite sulla base di una scheda condivisa contenente le 3 migliori pubblicazioni, le subject categories di competenza e l'indice h delle citazioni.

Nei mesi di febbraio e marzo 2012 i membri GEV hanno contattato personalmente i revisori proposti tramite una mail informale per sondarne la disponibilità. Nello stesso periodo il Presidente ha chiesto alle società scientifiche di Area 13 di proporre dei nominativi come revisori. In particolare ha chiesto alle società scientifiche di inviare un elenco di possibili revisori, con la richiesta esplicita di una quota significativa di revisori afferenti ad istituzioni straniere. In tale richiesta, ha raccomandato alle società di classificare le proposte di revisori tra le varie subject categories, in modo da assicurare una copertura di tutte le aree di ricerca. Tutte le proposte (derivanti dagli elenchi predisposti dal CINECA, dalle società scientifiche e dai singoli membri GEV) sono state raccolte in un'unica lista e successivamente approvate dal Presidente e dai coordinatori dei sub-GEV. In alcuni casi, sulla base del profilo scientifico del valutatore, le proposte dei membri GEV non sono state accolte, ed in ciascuno di questi casi il membro GEV è stato informato. Sono state attentamente considerate le proposte delle società scientifiche, ma queste ultime non sono state informate sull'esito delle proposte.

Grande attenzione è stata posta all'anonimato dei revisori, sia nella fase di predisposizione dell'elenco dei revisori, sia nella fase operativa della valutazione. I singoli membri GEV hanno potuto consultare, per scegliere i revisori, solo l'elenco dei revisori relativi alle Subject Categories di propria competenza.

La lista dei revisori è stata integrata durante il processo di valutazione sulla base delle necessità emerse a valle della trasmissione dei prodotti da parte delle Strutture, quando sono emerse aree di competenza poco coperte dai revisori inizialmente contattati, o la necessità di integrare il numero di revisori per subject categories in cui sono stati presentati molti prodotti o sostituire revisori non più disponibili o che si sono resi disponibili a valutare un numero limitato di prodotti.

Va peraltro notato che la previsione iniziale era di iniziare la peer review nel mese di giugno ed era pertanto stata richiesta la disponibilità dei revisori per il periodo giugno-ottobre. A causa dei ritardi nella trasmissione dei prodotti e dell'elaborazione dei dati da parte del CINECA, l'effettiva distribuzione dei prodotti è potuta iniziare solo a settembre, con la conseguenza che alcuni revisori non sono più stati disponibili.

Le integrazioni dell'elenco dei revisori sono state fatte nei mesi di novembre e dicembre, su proposta dei membri GEV e previa approvazione, caso per caso, da parte del Presidente e dei coordinatori sub-GEV, con la stessa procedura utilizzata per la composizione della lista di partenza.

La revisione peer review si è svolta dalla fine di settembre 2012 a febbraio 2013. Complessivamente, il GEV ha utilizzato 610 revisori esterni, distribuiti nelle subject categories indicati nella Tab. 1 (si noti che a molti revisori è stata attribuita anche una seconda subject category). Per ciascuno dei 3 sub-GEV quasi il 50% dei revisori è affiliato a un'istituzione straniera (Tab. 6); circa ¼ dei revisori lavora nel Regno Unito o negli Stati Uniti, e un restante quarto in altri paesi. Complessivamente l'80 percento dei revisori è di genere maschile.

La Tab. 7 riporta la distribuzione di h-index per i revisori; la media complessiva è 11,96, con valori leggermente più elevati per l'area statistica e leggermente inferiori per l'area aziendale.

La Tab. 8 indica che il numero medio di prodotti valutati è 17,50. Il numero è più elevato nell'area aziendale (23,4), che ha registrato anche il maggior numero di lavori in peer review, con conseguente carico di lavoro maggiore per i revisori. Il punteggio medio (normalizzato tra 0 e 1) attribuito dai revisori è di 0,41, con valori compresi tra 0,40 e 0,44 nei 3 sub-GEV. Anche la deviazione standard del punteggio attribuito dai revisori è simile nelle tre aree.

Tab. 6: Distribuzione dei valutatori esterni per sede dell'università o ente di ricerca e sub-GEV.

Sub-GEV	Country	Numero Revisori	%
Economia	**Italia**	148	0,53
	UK	34	0,12
	USA	39	0,14
	Altro	59	0,21
	Totale	**280**	**100,00**
Economia Aziendale e Finanza	**Italia**	117	0,55
	UK	32	0,15
	USA	22	0,10
	Altro	42	0,20
	Totale	**213**	**100,00**
Statistica e Matematica Applicata	**Italia**	64	0,55
	UK	11	0,09
	USA	13	0,11
	Altro	29	0,25
	Totale	**117**	**100,00**

Tab. 7: Distribuzione di h-index dei revisori esterni, per sub-GEV.

Sub-GEV	Media	sd	p10	p25	p50	p75	p90	N
Economia	12,29	7,94	5,00	7,00	10,50	15,00	22,00	280
Economia Aziendale e Finanza	9,93	7,18	3,00	5,00	8,00	13,00	21,00	213
Statistica e Matematica Applicata	14,83	7,88	7,00	9,00	14,00	20,00	26,00	117
Totale	**11,96**	**7,86**	**4,00**	**6,00**	**10,00**	**16,00**	**22,00**	**610**

Tab. 8: Caratteristiche dei valutatori esterni, numero di lavori valutati, media e s.d. dei lavori valutati, per sub-GEV.

Sub-GEV	Nazionalità Italiana	Sesso (Maschio)	Numero medio prodotti	Punteggio medio normalizzato	Dev. Standard Punteggio
Economia	0,53	0,83	14,44	0,42	0,04
Economia Aziendale e Finanza	0,55	0,77	23,42	0,40	0,03
Statistica e Matematica Applicata	0,55	0,77	14,07	0,44	0,04
Totale	**0,54**	**0,80**	**17,50**	**0,41**	**0,04**