

# Bibliometrics, what next ? Perspective of a journal editor



Gianluca Setti  
ENDIF, University of Ferrara  
gianluca.setti@unife.it

*stm*

STM Innovations Seminar, London,  
7<sup>th</sup> December 2012

# Aim of the Talk

- Overview on (citation-based) most successful bibliometric indicators from the point of view of an academic and past EiC of 2 IEEE journals (member/chair of several IEEE committees dealing with (quality of) publications)
  
- More specifically:
  1. Give an overview of pros and cons of various indicators
  2. Show that the "quality" of a journal as measured by bibliometric indicators is a multidimensional concept which cannot be captured by any single indicator (maybe one can use a "composed one")

# Bibliometric Indicators, i.e. ...numbers, numbers, numbers... 😊

Many bibliometric indicators exist, each aiming at measuring "journal quality": they should:

- 1. Give a result which corresponds to the technical quality of the papers published in that journal:** Science and the "Journal of Obscurity" should have a very different indicator
- 2. Be "fair" if applied to different areas:** different areas/communities may have different citations practice (long/short citation list)
- 3. Be immune to external manipulation:** it should be very difficult to artificially manipulate its value

Note: Use of citations to measure quality is not perfect ... and this is where altmetrics can be of great use

# IF and its critics...

$$\text{IF}_i = \frac{C_{\rightarrow i}^{\Delta_1 \Delta_2}}{p_i^{\Delta_1}} \quad \Delta_2 = \{Y_n\}$$
$$\Delta_1 = \{Y_{n-1}, Y_{n-2}\}$$

- **Pros:** simple, easy to compute, known and diffused
- **Cons/critics:**
  1.  $\Delta_1$  of 2 years only to account for citations may not be enough in some areas to mature citation peak  $\Rightarrow$  **IF varies very significantly among (sub)areas**
  2. Citations are counted in the same way **independently of the source** (i.e. a citation obtained from *Science* is the same as the “*Journal of Obscurity*”)
  3. IF is **liable to active manipulation:**
    - "coercive" self-citations
    - citations to notes/letters count at the numerator but notes/letters are not in  $p_i^{\Delta_1}$

# Coercion and Self-Citations level

## Coercive Citation in Academic Publishing

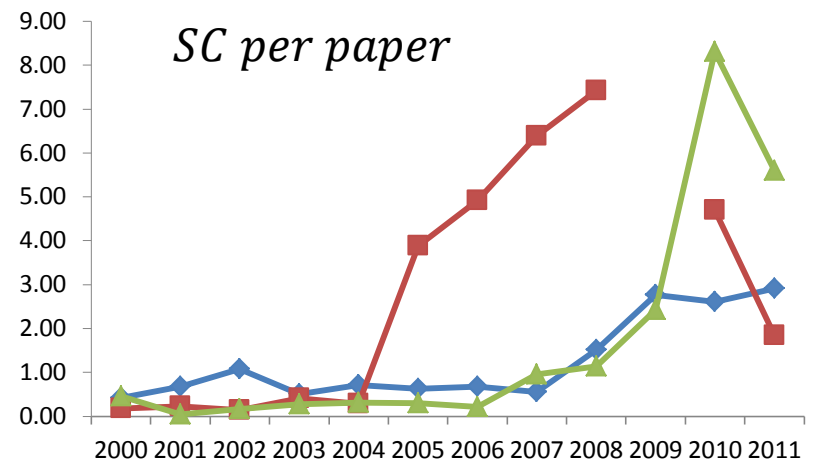
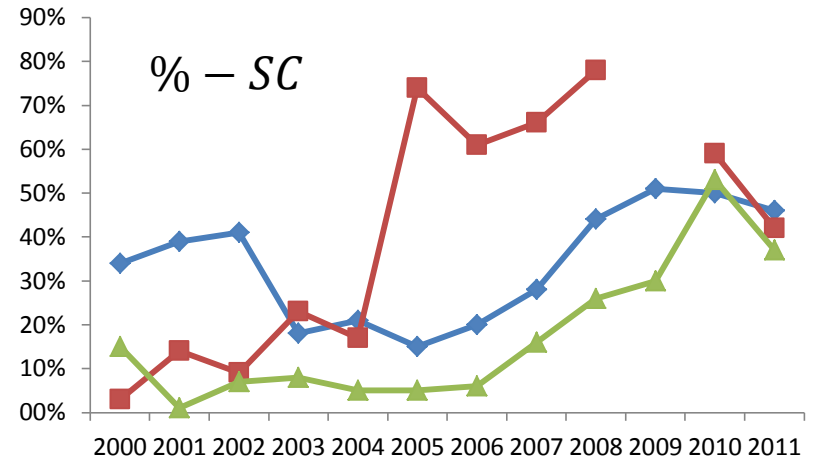
Allen W. Wilhite\*† and Eric A. Fong\*

3 FEBRUARY 2012 VOL 335 SCIENCE www.sciencemag.org  
Published by AAAS

- EICs of 175/832 journals in the area of economics, sociology, psychology, and multiple business disciplines were found to "coerce" self-cites
- Coercing was more frequent with young authors than experienced ones
- Relation to area: if one coerces others will most likely follow
- For profit published were found more prone to coerce than "academic" ones (more interested to maintain their reputation)

And in other areas?

1. **Laser and Particles Beams** (Phy Applied)
2. **Cortex** (Neuroscience)
3. **Int. Journal of Hydrogen Energy** (Energy and Fuels)



# Why coercing?

- There are (unfortunately) unintended use of the IF:
  1. Evaluation of a **single** paper in a journal: journal indicators are average quantities and give therefore **no indication** of the quality of a single paper published in it)

Since the distribution of citations of the papers in a journal is extremely skewed assuming that paper  $P_{JA}$  is better than paper  $P_{JB}$  since  $IF_{JA} > IF_{JB}$  may be **more often false than true**

2. Salary increase: Chinese government pays scientists for publication in high IF journal\*. This is, at best, measuring quality in one dimension (the IF) and always provide "distortion"...
3. Evaluation of the CV of a single scientist (for tenure/promotion/grant assignment): sum of IF, average of IF...

The unintended use of the IF **made it the target and not the measure** and “when a measure becomes a target, it ceases to be a good measure” -  
Goodhart’s law

(from D. Arnold, K. Fowler, "Nefarious Numbers", Notices of the AMS, vol 58, n.3, pp 434-437)

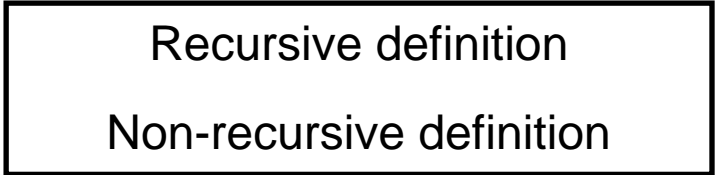
# To solve IF issues...

- 8 "successful" new indicators (5 in either WoS or Scopus)

Property/Feature	Bibliometric Index									
	IF	Popularity Measures			Prestige Measures			AI	SJR	
Publication Window	2 years	5 years	1-4 years	3 years	5 years	5 years	5 years	5 years	3 years	
Citation Window	1 year	1 year	1-4 years	1 year	1 year	1 year	1 year	1 year	1 year	
Self-Citations	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes (max 33%)	
Subject/Field Normalization	No	No	Yes (wrt average citations received by a prescribed subject field)	Yes (wrt citations from set of citing journals)	Yes (wrt average citations from any set of journals)	Not needed	Not needed	Not needed	Not needed	
Cited/Citing-side Normalization	N/A	N/A	Cited	Citing	Citing	Yes (based on influence redistribution)	Yes (based on prestige redistribution)	Yes (based on prestige redistribution)	Yes (based on prestige redistribution)	
Document Type Numerator	Articles, letters notes, reviews	Articles, letters notes, reviews	Articles, letters notes, reviews	Articles, reviews	Not mentioned explicitly	Not mentioned explicitly	Articles, letters notes, reviews	Articles, letters notes, reviews	Articles, reviews	
Document Type Denominator	Articles, reviews	Articles, reviews	Articles, letters notes, reviews	Articles, reviews	Not mentioned explicitly	Not mentioned explicitly	Articles, letters reviews	Articles, letters reviews	Articles reviews	
Weight of Citing Source	None	None	None	None	Yes	Yes	Yes	Yes	Yes	
Data Base	JCR	JCR	(WoS)	Scopus	(JCR)	(JCR)	JCR	JCR	Scopus	

- Increase in citation window  $\Delta_1$ :** 3 or 5 years
- Introduction of subject field normalization:** explicit (JFIS, SNIP, AF) or by construction (IW, EF, AI, SJR)
- Do not consider self-cites or (even better do so till x%):** eliminate the inflation issue
- Consider the same** (apart in some cases from "notes") **kind of documents both at numerator and denominator:** eliminate another cause of inflation

Popularity vs Prestige



# Recursive (Pagerank) Prestige Measures - I

- The **EigenFactor** is computed by **ISI-Thomson** using the “same” algorithm used by Google to rank web pages
- How  $EF_i$  (for journal  $k$ ) is “roughly” defined?:

"self-citations" are not included

$$\pi_i[k] = (1 - \alpha) \frac{p_i^{\Delta_2}}{\sum_{k=1}^N p_k^{\Delta_2}} + \alpha \sum_{j=1, j \neq i}^N \frac{c_{ji}^{\Delta_2 \Delta_1}}{\sum_{l=1, l \neq j}^N c_{jl}^{\Delta_2 \Delta_1}} \pi_j[k-1]$$

$$\sum_{i=1}^N \pi_i[k] = 1$$

### Remarks

1. The more the journal  $j$  is "important" ( $\pi_j$  is large) the more a citation from it to journal  $i$  increases  $EF_i$
2. all citations given by journal  $j$  (normalization by citation potential)
3. The  $EF_i$  represents the probability that a random reader picking journals at random an following citation will eventually read journal  $i$
4. The damping factor  $\alpha$  takes into account that the reader will at some point stop reading (usually 0.85) and "begin reading again" proportionally to the "journal size"

- One needs to compute

$$\pi_i = \lim_{k \rightarrow \infty} \pi_i[k]$$

$$EF_i = 100 \sum_{j=1}^N \frac{\pi_j}{C_{\leftarrow j}^{\Delta_2 \Delta_1}} c_{ji}^{\Delta_2 \Delta_1}$$



# Recursive (Pagerank) Prestige Measures - II

- The **Article Influence** is roughly the EF “normalized to” the number of papers published by each journal (similar “physical meaning” w.r.t. IF) + a normalization to have AI=1 for the median journal

$$AI_i = \beta \frac{EF_i}{p_i^{\Delta_2}}$$

- **Pros (EF/AI):**
  1. Citations **are now weighted** depending on the source (a citation from *Science* is valued more than one from the “*Journal of Obscurity*”)
  2. Time window for computing citations ( $\Delta_2$ ) **is 5 years**. This index are expected to exhibit less fluctuations over time
  3. Journal self-citations **are not** considered. The index is less prone to “external influence”
- **Cons (EF/AI):**
  1. **Not necessarily correct** to eliminate **all** self-cites. SJR has a “similar” definition wrt EF/AI but: adopts self-cites till 33%
  2. More difficult to understand and compute

# Other indicators and their relationships

## Popularity:

1. **5 Year Impact Factor (5YIF):** IF with  $\Delta_2$  of 5 years
2. **Journal to Field Impact Source (JFIS):** it is basically an IF with the citing window and citation window equal (from 1 to 4 years) and with a normalization wrt the "same quantity" in the area/category (the value is computed a sum of the averages for "articles", "reviews", "notes" and "letters" which account for different kind of journals)
3. **Source Normalized Impact per Paper (SNIP):** ratio of "an IF" over a citation window of 3 years and the "relative database citation potential of the journal" i.e. the average number of citations contained in any journal citing journal  $i$  (normalized to have median = 1)

## Prestige:

1. **Influence Weight (IW):** it is again an "eigeninfluence" (recurrent) measure, with a different normalization (Idea from 1976, inspired Google PR)
2. **Audience Factor:** Basically an IF but there is weight for the citation depending on the average number of citations given wrt to the average of the area

**Note:** if  $\alpha = 0$ ,  $AI_i \propto AF_i$  and if  $\alpha = 1$   $AI_i \propto IW_i$

# Using the h-factor

- Proposed in 2006 by Glaenzel et al. and already computed by f.i. Scopus
- Different way to do this (h from birth, h from a given year on, h in a fixed time window, h-core, h-median)

- Google started using **h-5** **h-5-median**, since April 1, 2012
- In **top 100** "journals" there is also ArXiv or RePEc

Google scholar    [My Citations](#)

Top publications in: [English](#) - [Chinese](#) - [Portuguese](#) - [German](#) - [Spanish](#) - [French](#) - [Italian](#) - [Japanese](#) - [Dutch](#) - [Korean](#) [Learn more](#)

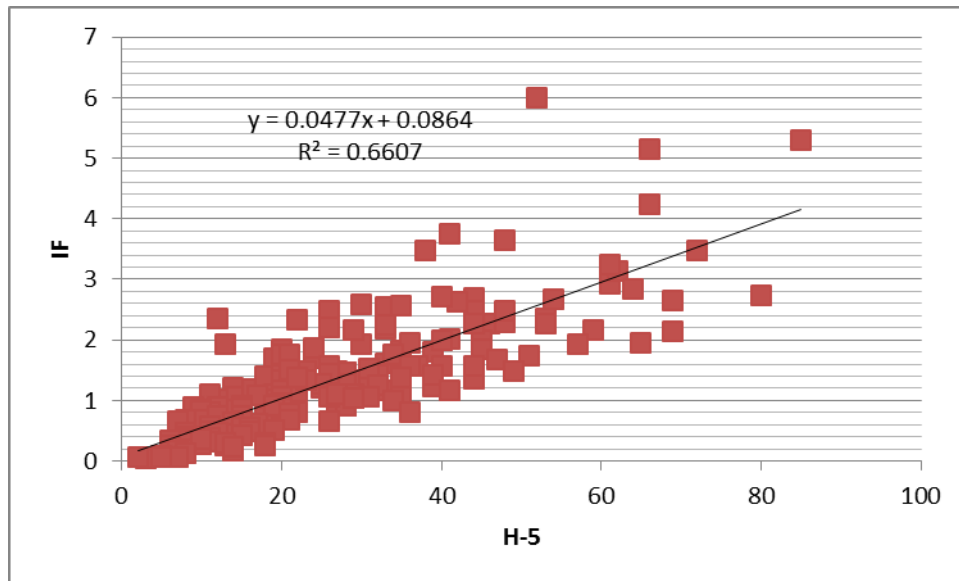
Title	h5-index	h5-median
1. Nature	295	427
2. New England Journal of Medicine	274	450
3. Science	265	388
4. RePEc	259	356
5. arXiv	256	367
6. The Lancet	205	313
7. Social Science Research Network	205	290
8. Cell	195	279
9. Proceedings of the National Academy of Sciences	189	237
10. Nature Genetics	174	268
11. Journal of Clinical Oncology	173	229
12. JAMA: The Journal of the American Medical Association	171	246
13. Physical Review Letters	162	213
14. Circulation	159	251
15. Chemical reviews	144	248
16. Blood	141	192
17. The Astrophysical Journal	140	181
18. Journal of the American College of Cardiology	139	192
19. Journal of the American Chemical Society	138	174
20. Nucleic Acids Research	135	239
21. Angewandte Chemie International Edition	133	190
22. Nature Reviews Molecular Cell Biology	128	219
23. Nature Medicine	127	191

# Using the h-factor

- Proposed in 2006 by Glaenzel et al. and already computed by I.I. Scopus
- Different way to do this (h from birth, h from a given year on, h in a fixed time window, h-core, h-median)
- Google started using **h-5**  
**h-5-median**, since April 1, 2012
- In **top 100** "journals" there is also ArXiv or RePEc
- **Remarks**
  1. Frozen picture in time (April 1, first, now Nov 15) with some source of citation eliminated to be "sure" it is an actual "publication" (**moderates some critics on Scholar**)
  2. Popularity and usage could increase due to accessibility

# What do we do with all these data??

- Main problem in bibliometrics is that "true quality" **cannot be measured** (maybe not even defined).
- We have only "**indirect observables of quality**" and we can only correlate them, by computing Pearson CC, Spearman ranking CC, Kendall ranking CC (more though...)
- Example: Engineering, E&E 2010 (IF, 5YIF, EF, AI, h-5), only journals having all of them



Pearson CC					
	IF	5YIF	EF	AI	h5
IF	1	0.959	0.628	0.896	0.813
5YIF	0.959	1	0.639	0.949	0.848
EF	0.628	0.639	1	0.646	0.828
AI	0.896	0.949	0.646	1	0.781
h5	0.896	0.848	0.828	0.781	1

# What do we do with all these data??

- Main problem in bibliometrics is that "true quality" **cannot be measured** (maybe not even defined).
- We have only "indirect observables of quality" and we can only correlate them, by computing Pearson CC, Spearman ranking CC, Kendall ranking CC (more though...)
- Example: Engineering, E&E 2010 (IF, 5YIF, EF, AI, h-5), only journals having all of them

	Spearman Rank CC				
	IF	5YIF	EF	AI	h5
IF	1	0.966	0.749	0.904	0.852
5YIF	0.966	1	0.757	0.952	0.875
EF	0.749	0.757	1	0.753	0.896
AI	0.904	0.952	0.753	1	0.832
h5	0.852	0.875	0.896	0.832	1

	Kendall Rank CC				
	IF	5YIF	EF	AI	h5
IF	1	0.847	0.559	0.737	0.671
5YIF	0.847	1	0.566	0.823	0.697
EF	0.559	0.566	1	0.565	0.720
AI	0.737	0.823	0.565	1	0.648
h5	0.671	0.697	0.720	0.648	1

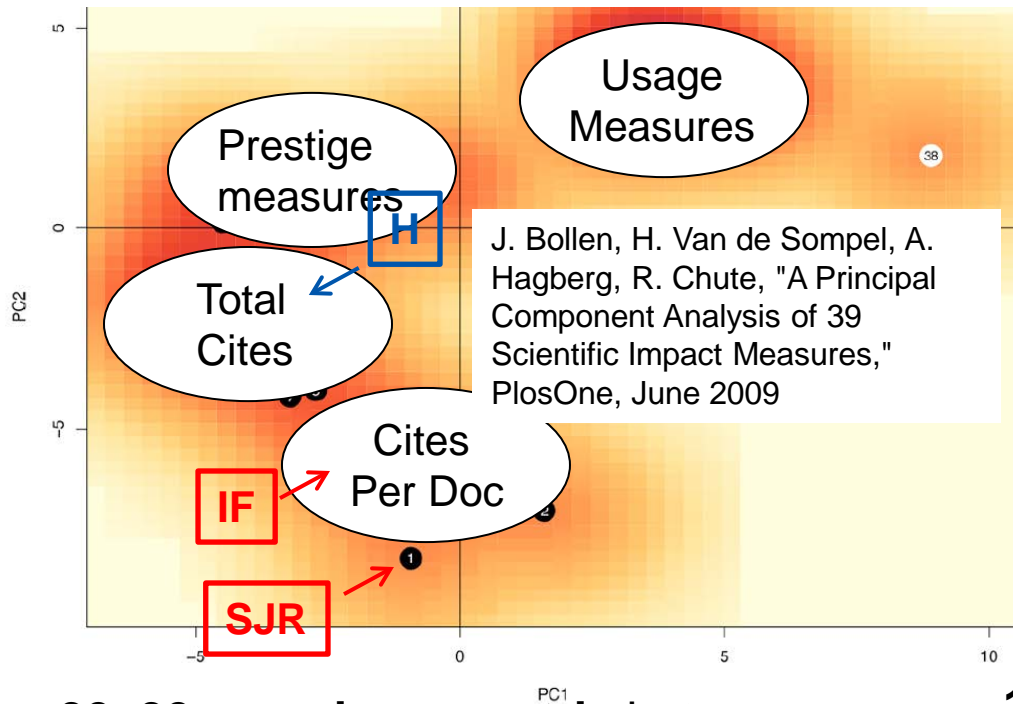
# What do we do with all these data??

- Main problem in bibliometrics is that "true quality" **cannot be measured** (maybe not even defined).
- We have only "**indirect observables of quality**" and we can only correlate them, by computing *Pearson CC*, *Spearman ranking CC*, *Kendall ranking CC* (more though...)
- Example: Engineering, E&E 2010 (IF, 5YIF, EF, AI, h-5), only journals having all of them

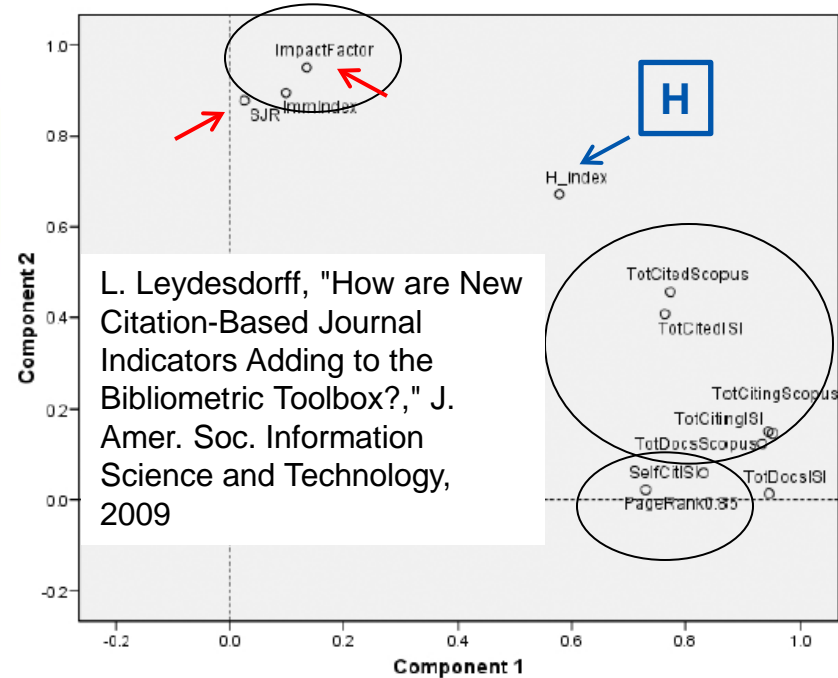
## Remarks:

1. All measures are positively correlated (kind of "minimum requirement")
2. IF, 5YIF, AI correlated very well (obvious for the first 2)
3. EF correlated less strongly with IF, 5YIF, AI (EF is a per-journal measure)
4. H5 correlates "on the average" with all others (somehow "in between" the 2 other groups)

# Two PCA Analysis of bibliometric indicators



**39x39 covariance matrix** between indexes computed using Scimago, 2007 JCR and MESUR project for usage



**13x13 covariance matrix** between indexes computed using Scimago and 2007 JCR (no usage)

Compute the "principal components":

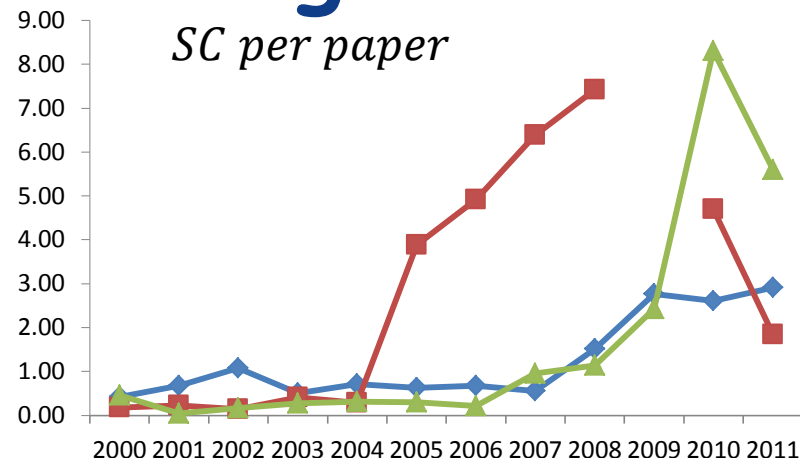
1. the problem is mainly 2-dimentional
2. Different clusters are present: prestige, popularity and cites **measure different aspects of quality**
3. One **cannot use only one indicator** to "measure journal quality"



# What kind of info can we get?

- Back to coercion problem:

With respect to "SC per paper"  
*Cortex* in 2009 is worse than *Laser and Particles Beams*. Why was not removed?



	Laser and Particles Beams (Physics, Applied)					Cortex (Behavioral Science)				
	2007	2008	2009	2010	2011	2007	2008	2009	2010	2011
Rk-IF	6	8	SUP	17	49	8	19	7	4	4
Rk-EF	32	54	SUP	45	57	9	15	12	15	10
Rk-AI	50	67	SUP	70	69	15	21	17	16	14

Using more than one indicator may help in taking appropriate decisions  
 (one could even think to use a linear predictor to detect anomalies!)

# A "Level-2" Indicator

- Multiple indicators must be used to assess publication quality
- What if one needs to have a single indicator to rank (categorize) publications in a certain area?

Use a level-2 composed indicator by using a Principal Component Analysis (PCA)

1. Expresses each indicator in standardized form (Z-variable, zero mean and unitary variance)
2. Combine the standardized indicator by using PCA (maximize variance of data projection, i.e. "information")

Example: Engineering, E&E 2010 (only journals having indicators)\*

Pearson CC	IF	5YIF	EF	AI	h5
IF	1	0.959	0.628	0.896	0.813
5YIF	0.959	1	0.639	0.949	0.848
EF	0.628	0.639	1	0.646	0.828
AI	0.896	0.949	0.646	1	0.781
h5	0.896	0.848	0.828	0.781	1

# A "Level-2" Indicator

- Multiple indicators must be used to assess publication quality
- What if one needs to have a single indicator to rank (categorize) publications in a certain area?

Use a level-2 composed indicator by using a Principal Component Analysis (PCA)

1. Expresses each indicator in standardized form (Z-variable, zero mean and unitary variance)
2. Combine the standardized indicator by using PCA (maximize variance of data projection, i.e. "information")

Example: Engineering, E&E 2010 (only journals having indicators)\*

Pearson CC	IF	5YIF	EF	AI	h5	PCA
IF	1	0.959	0.628	0.896	0.813	0.955
5YIF	0.959	1	0.639	0.949	0.848	0.973
EF	0.628	0.639	1	0.646	0.828	0.780
AI	0.896	0.949	0.646	1	0.781	0.957
h5	0.896	0.848	0.828	0.781	1	0.887
PCA	0.955	0.973	0.780	0.957	0.887	1

# A "Level-2" Indicator

- Multiple indicators must be used to assess publication quality
- What if one needs to have a single indicator to rank (categorize) publications in a certain area?

Use a level-2 composed indicator by using a Principal Component Analysis (PCA)

1. Expresses each indicator in standardized form (Z-variable, zero mean and unitary variance)
2. Combine the standardized indicator by using PCA (maximize variance of data projection, i.e. "information")

Example: Engineering, E&E 2010 (only journals having indicators)\*

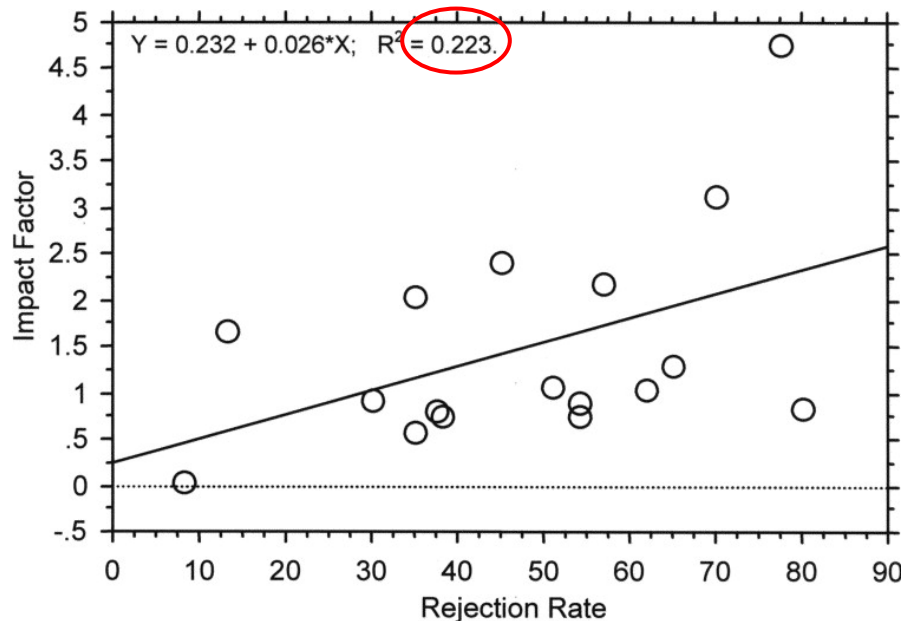
## Remarks:

1. PCA is "more correlated" with other indicators than h-5 apart for EF
2. Could be a valid alternative as a "single indicator"
3. Used in the Italian Research Evaluation Exercise VQR2004-2010

# A final word of caution

- There is no study ensuring that bibliometric index "measure scientific quality". One assumes, f.i. that a "high IF" is also a "selective journal"

...



A. Kurmin, T. Krimis, "Exploring the Relationship Between Impact Factor and Manuscript Rejection Rates in Radiologic Journals, Acad Radiol 2006; 13:77–83

- Data for journals in other areas and more recent years (2008) show a similar trend (also for EF and AI!)

# Conclusive Remarks

Bibliometric indicators exist, each aiming at measuring "journal quality" and they measure "quality" in a different way

1. One should **not use a single indicator** (IF, but any other else as well) to measure journal impact
2. Using more indicators can also give more information on possible manipulations of bibliometric indicators
3. Ranking could possibly be determined using a Level-2 indicator
4. Bibliometric indices are very useful tools to **evaluate journal impact**, but **cannot be exploited (alone) to evaluate the quality of single papers or single scientists**

Is altmetrics the solution?