Citation for published version

Mingers, John (2009) Measuring the Research Contribution of Management Academics using the Hirsch-Index. Journal of the Operational Research Society, 60 (9). pp. 1143-1153. ISSN 0160-5682.

DOI

https://doi.org/10.1057/jors.2008.94

Link to record in KAR

http://kar.kent.ac.uk/22292/

Document Version

Author's Accepted Manuscript

Measuring the research contribution of management academics using the Hirsch-index

John Mingers

Kent Business School, University of Kent, Canterbury CT7 2PE, UK

phone: 01227 824008, e-mail: j.mingers@kent.ac.uk

V2.1, June 2008

Submitted to JORS

Measuring the research contribution of management academics using the Hirsch-index

Abstract

There is an increasing emphasis on the use of metrics for assessing the research contribution of academics, departments, journals or conferences. Contribution has two dimensions: quantity which can be measured by number/size of the outputs, and quality which is most easily measured by the number of citations. Recently, Hirsch proposed a new metric which is simple, combines both quality and quantity in one number, and is robust to measurement problems. This paper applies the h-index to three groups of management academics – BAM Fellows, INFORMS Fellows, and members of COPIOR – in order to evaluate the extent to which the h-index would serve as a reliable measure of the contribution of researchers in the management field.

Keywords: citations, Hirsch, h-index, metrics, productivity, RAE, research

Introduction

Measuring the research contribution of academics is of increasing importance in the management of universities. This is especially the case in the UK where the long-standing Research Assessment Exercise (RAE), which has been based entirely on peer-review, is to be supplanted or replaced by a metrics-based process (DfES, 2006) that will be known as the Research Excellence Framework (REF). The RAE is a national system, carried out every few years at great expense and time, that aims to evaluate the research quality of all departments in all UK universities. The original purpose was to inform decisions about where research funding should be allocated. It has become, however, the most important way of ranking universities in terms of their research.

There may be many academics who condemn the idea that a person's contribution can be reduced to one or two numbers (Kelly and Jennions, 2006; Macdonald and Kam, 2007). I have some sympathy with that view, but the stance in this paper is that if metrication is going to occur, as seems inevitable, then it is important that research is carried out so that the most reliable and effective measures are used. In this case I report on an investigation into a new measure proposed recently by Hirsch (2005) which has attracted much attention.

There are a variety of reasons and contexts for evaluating research contribution and quantitative measures seem to be becoming more important in all of them. For an individual researcher they can inform decisions about hiring, tenure, promotion or the extent of teaching commitment; for a research group the questions may concern funding, relative strength as in the RAE, or continued existence; for journals their relative standing in important in terms of circulation and quality of submissions.

In principle, research contribution has two components: quantity and quality. Quantity is relatively easy to measure by number of outputs or, perhaps in a humanities discipline where it is mainly books, by number and length. Quality is much harder to measure and the main, non-peer review, approach is to evaluate the impact of the work through the number of citations received. The most well known citation measure, which is used to compare journals, is the impact factor (Garfield, 1972) which is based on the number of citations per paper received by the journal in the last two years. There are known problems of using citations at all (MacRoberts and MacRoberts, 1987; Seglen, 1997; Jennings, 1998; Brumback, 2008), but nevertheless this seems to be the main way forward at the moment.

This paper explores a particular citation measure - the h-index – which has been proposed recently and has generated much interest in the information science and bibliometrics field. Hirsch (2005) and Glänzel (2006b) have discussed its main characteristics. This exploration is done within the business and management context by examining three different groups of senior management academics: UK operational researchers, US operational researchers, and UK general management academics. The first section discusses the literature concerning the h-index and alternative measures; the second the methodology of the study; and the third the results. It goes beyond the individual researcher to consider applying the h-index to research groups and to journals.

Citation Measures and the Hirsch-Index

Citation measures and criteria for comparing them

The most obvious citation measures that have been used are either the total number of citations generated by an individual or group, or some form of mean citations per paper (or other research output). van Raan (2003) discusses the main measures that have been used: P, the number of publications, C the total number of citations, CPP the mean citations per paper and %Pnc, the percentage of papers never cited. It is the case that disciplines vary immensely in their citation practices and so it can be useful to normalise the mean rate to the overall mean rate of citations in the actual set of journals used, or within the research field as a whole. The data used by van Raan and the Centre for Science and Technology Studies is drawn exclusively from the ISI Web of Science which gives only limited coverage in many disciplines, a point discussed further below.

Given a range of measures, including the h-index, how should we evaluate them? I would put forward three general criteria: *effectiveness* in that it does measure research contribution; *reliability* in that it is easy to measure and robust to data problems (which are many in this area as we shall see); *transparency* in that it is easy to understand and interpret.

In terms of effectiveness, I suggest that research contribution is a combination of productivity, i.e., number of publications, and international impact or influence (van Raan, 2003). If this is accepted then there are problems with the traditional measures of total citations or mean citations, even in the sophisticated van Raan form, as the following (somewhat extreme) example shows. Researcher A published a very influential paper ten years ago (100 citations) but has not published anything since; researcher B publishes many papers but they are repetitive and inconsequential gaining few citations (100 with 40 papers); researcher C is productive and his or her publications are seen as significant (100 citations with 10 papers). A has citations but little productivity, B has productivity but no impact, while I would argue that C makes the best contribution with both quantity and quality. Neither total citations nor mean citations reflects this. Total citations are the same for each and so do not distinguish between them. Moreover, as we can see, the total number of citations can be distorted by a single "big hit". The mean citations clearly favours those with fewer publications. Indeed, if academics knew that their work was going to be measured in this way in the future one could argue that it might be in their interest to reduce the number of papers they published – surely a counter-intuitive result.

The h-index

The h-index overcomes this problem by combining both quantity and impact in one measure. The index is defined as:

> "a scientist has index h if h of his/her N papers have at least h citations each and the other (N-h) papers have no more than h citations each" (Hirsch, 2005,p. 16569).

Thus an index of 20 (which is quite high in social science) means that the person has 20 publications each of which has *at least* 20 citations.

We can illustrate strengths and weaknesses with some examples:

Table 1 about here

Table 1 shows the publications of three actual academics with about the same length of service (W, X, Y) and a made-up one, together with some citation measures. W has produced lots of papers but is only moderately well cited; X has produced less but has same number of citations and two highly cited papers; and Y has produced only a small number but is the most highly cited. These examples illustrate the trade-off between quality and quantity. If we are primarily interested in quantity then W is best, especially compared with Y, but in terms of quality Y is best with many more total citations.

If we look to find a balance then I would argue that the mean number of citations is very misleading. According to that, Y is three times better than W and twice as good as X which I think is unrepresentative and could motivate researchers *not* to produce papers. The h-index is much more balanced giving W and Y the same score of 5, slightly better than X which, I suggest, is a sensible result.

We can also see that the h-index, like any single measure, has its own peculiarities some of which are addressed in more detail later. First, gaining additional citations for papers that are already above h or are well below h has no effect on h. It is only those just below h at any point in time that can change it by getting above the h threshold. So for W one extra citation for papers 6 or 7 would increase h to 6. This may seem like a strange effect but in the long-term the value of h does rise in a consistent manner. Second, producing additional papers of itself does not increase h. This only happens when the paper gets more citations than the current value of h. But conversely, having only a small number of papers does put an immediate upper limit on h.

Third, example Z shows a rather perverse effect. Most people would say that Y is clearly better than Z – they have more papers, more citations, and a more highly cited paper – yet their h-index is the same. This is a consequence of its insensitivity to total number of publications, and highly cited papers. It is interesting, however, that this also occurs with the mean citation measure. In fact, the mean actually shows Z to be better than Y!

The primary advantages and disadvantages of the h-index are:

- It measures both productivity and influence or impact of the papers as opposed to other common measures such as: total number of papers which only measures productivity; total number of citations which is hard to measure and can be inflated by a few "big hits"; citations per paper which penalises high productivity; or number of "significant papers" which is somewhat arbitrary.
- The h-index is simple to calculate, easily understood and very robust to the difficulties of measuring citations since it only uses the few, most highly cited, papers.
- It can be applied at several levels of aggregation – individual, research group, journal, or department.
- Any type of research output can be included and it is not affected by outputs with zero citations
- The h-index correlates well with other standard bibliometric measures (Bornman and Daniel, 2005; van Raan, 2005).

There are recognised disadvantages, some of which have led to the development of alternative measures:

- The h-index for an individual can only ever rise, even after they stop publishing. It is directly proportional the length of a person's career. This puts early career researchers at a disadvantage (Burrell, 2007b) and does not properly reflect a fall-off in productivity in later years (Sidiropoulos et al., 2006).
- The fact that the index is not influenced by very highly cited papers could be seen as a disadvantage. Given two people with a similar index, one could have papers with ten times more citations than the other but this would make no difference (Egghe, 2006).
- The index is clearly very different across disciplines as it reflects the publication patterns of the discipline.
- The index does not take into account patterns of co-authorship. These certainly differ between disciplines, but can also differ between countries within a discipline (Batista et al., 2006).
- It is only a single measure and therefore cannot account for all aspects of research performance (van Raan, 2005).

Several authors have calculated h-indices for groups of senior academics in different disciplines. Hirsch (2005) found that some prominent physicists ranged from 110 (E. Witten) to 62 (S. Hawkins); that Nobel prize-winning physicists were from 79 to 22; and that the top ten in the life sciences ranged from 191 to 120. These are clearly huge figures and reflect the publication habits of the natural sciences. In contrast, Cronin and Meho (2006) found that 31 US information scientists ranged from 20 to 5; Oppenheim (2007) measured UK information scientists from 31 to 5; Saad (2006) found consumer researchers ranged from 17 to 3; and Sidiropoulos et al (2006) measured computer scientists between 24 and 14. Clearly, the h-index is very sensitive to the disciplinary background and Iglesias and Pecharroman (2006) have calculated a set of scaling factors to correct an h-index relative to the discipline of physics. To give some examples, molecular biology is 0.44, computer science is 1.75 and social science is 1.60.

There has also been some theoretical investigations of the properties of the h-index. Glänzel (2006a) developed two models assuming that the cumulative citation process followed either the Paretian distribution or the Price distribution. This was not empirically verified. The main results were:

- the h-index is proportional to the (α+1)th root of the number of publications (P), where α is the "characteristic tail" parameter of the distribution (Glänzel and Schubert, 1988). The Price distribution has $\alpha=1$ and for $\alpha > 2$ the distribution is Gaussian. So h would typically be the square or cube root of the total number of publications. Van Raan (2005), in an evaluation of 147 chemistry research groups (not individuals), found that $h = 0.73P^{0.52}$.
- The number of citations of the h papers included in the h-index is proportional to the square of h. This is almost obvious as h papers each having at least h citations must have at least h^2 citations.
- The highest cited publication is also a power function of h.

These results relate h to the quantity of output, while Burrell (2007a) has considered the behaviour of h over time. His stochastic model assumes that an author publishes papers each year according to a Poisson process (parameter θ); that each paper generates citations also according to a Poisson process but where the rate varies from paper to paper; and that this citation rate varies as a gamma distribution (shape ν, scale α). Under these assumptions the distribution of the number of citations for a randomly chosen paper follows the beta distribution and the distribution of the number of papers receiving at least n citations by time T can be deduced. This enables estimates of the expected theoretical value of h to be made. These depend on four parameters – publication rate θ; gamma parameters ν and α, where ν/α is the author's average citation rate across papers, and the length of the publishing career T.

The main results are:

- that h is almost directly proportional to the career length, T which means that it is biased towards long-standing, but not necessarily highly productive, researchers;
- that h is approximately a linear function of the log of the author's publication rate, θ;
- that h is approximately a linear function of the log of the author's mean citation rate, ν/α.

Again, these are theoretical results that have not been empirically validated but they seem intuitively reasonable.

In response to some of the limitations of the h-index, variations have been proposed. Some of these will be illustrated in our results. Egghe (2006) suggested the g-index as a way of being more responsive to authors with very high citations of their top publications. "A set of papers has a g-index g if g is the highest rank such that the top g papers have, together, at least g^2 citations" (p. 132). Thus a g-index of 20 implies that the top 20 publications have a total number of citations of at least 400. Papers which are very highly cited will push up the g-index while not affecting the h-index.

Sidiropoulos et al (2006) addressed the issue that the h-index was biased towards those later in their career, who may even have stopped publishing. Their "contemporary" h-index weights the citations in terms of how recent the cited paper is. In their particular experiment, a citation of a paper published in the current year would have a weight of 4; papers 4 years old would have a weight of 1; and papers 10 years old would have a weight of 4/10. They also produced models to account for the age of the citation, and the number of papers published.

Burrell (2007b), whose model we discussed above, has suggested that the rate of development of the h-index should be used to overcome the bias towards long-serving academics. This can be done either by dividing the h-index by the career length of the academic this giving a raw h-rate per year or, if the time series data is available (e.g., as in Liang (2006)) then by performing a regression against time. He finds that this does make a significant difference to the rank ordering in some cases, even with senior academics.

Methodology

This study estimated h-indices, and other statistics, for three groups of senior academics in the management field:

- Members of the Committee of Professors in OR (COPIOR). This is a UK organisation open (after election) to Professors whose work is relevant to operational research. There are around 50 members although some are retired.
- Fellows of the British Academy of Management (BAM). These are generally UK academics elected as Fellows for their significant contribution to the field of management. There are around 50.
- Fellows of the Institute for Operations Research and the Management Sciences (INFORMS). These are primarily US academics elected for their significant contribution to operational research. There are around 120.

This enabled comparisons to be carried out between OR academics in the UK and the US, and between UK OR academics and UK academics in other management disciplines. In each of the three groups a random sample of 30 people was selected. Thus the overall results in Tables 2 – 5 apply only to the sample not the three groups as a whole.

The main effort was put into accurately collecting statistics on publications and citations for these 90 people but, as became clear, there are many problems in producing this data.

Source of data

The first issue is the source of data. Traditionally, the ISI database Web of Science (WoS) has been the main source of citation statistics but there are alternatives such as Google Scholar (GS) or more subject specialised databases such as the DBLP digital library. WoS and GS each have particular advantages and disadvantages (Harzing, 2007b; Walters, 2007): broadly WoS is the more rigorous but is limited to citations in journals that are included in the ISI database. This excludes some journals, but also books, conference proceedings, and dissertations. It can therefore significantly underestimate an academics publications and citations. Meho and Yang (2007) compared the citations for 1000 papers in information science published between 1996 and 2005 and found only a 30% overlap between GS and the others, largely because GS produced twice as many citations.

GS also has limitations: it can sometimes include non-scholarly citations such as handbooks or library guides; it too does not include all journals; it is less good for older publications; and its automatic processing can produce nonsensical results. But studies (Saad, 2006; Meho and Yang, 2007) have shown that when compared with WoS in terms of ranking researchers the results are very similar. Because the h-index primarily measures the amount of output and citations, it was decided to use GS because of its greater coverage. In fact, GS was not used directly – it was accessed through the *Publish or Perish* (PoP) software (Harzing, 2007a) specially designed to be used for citation analysis with GS.

The citation data covered the full length of time available in Google Scholar – up to 49 years in one case since the first publication. In fact, in any revised RAE it is likely that a much shorter time frame will be used, perhaps only 5 years. This would clearly have significant effects on any citation measure. There is much evidence that citation behaviour differs significantly between disciplines. In the hard sciences citation numbers peak relatively quickly, but in social science the peak is often around ten years with papers being cited for 20 or 30 years (Mingers, 2008). This issue needs further research.

Problems of data production

Whichever data source it used, there are immense problems in producing accurate counts of the total number of publications and citations for individual authors, particularly but not exclusively where they have common names.

1. For an author with an uncommon name, especially when they use more than one initial, the list of publications that PoP generates is reasonably accurate subject to points 2-3 below. However, with a common name and only a single initial it sometimes proved impossible to accurately identify publications and citations.
   GS produces a maximum of 1000 results for any query. This limit is often reached for researchers with common names as hundreds of academics may share the name. Steps can be taken to try to eliminate the spurious ones. PoP allows you to exclude incorrect names. This works if the author uses two initials as all others can be excluded but does not if there is only one initial. PoP also allows you to exclude particular general subject areas (e.g., medicine or engineering) but this often produces false results as genuine papers or citations are excluded. This appears to be a problem with the GS classification. To give an example, consider one of the academics who works in a cross-disciplinary area, with a common surname, who only uses one initial. Their departmental website lists 28 publications including conferences and book chapters; WoS (in which you can specify institution) lists 21 papers in the ISI journals. GS begins by listing 997 papers (in reality constrained by the 1000 limit) with 64,732 citations for a very large number of people with that surname. This cannot easily be reduced by excluding names as only one initial is used. Most subject categories have to remain because she could be cited in social science, computer science, life science, medicine and biology. So the

only way to proceed would be to have an exhaustive list of publications (something not generally available) and go through manually.

2. Citations often mis-cite a paper. Thus the same paper may appear several times with slightly different details and it is not always easy to decide if they are in fact the same paper. This inflates the number of papers. Looking at myself, where I know the publications, nearly a third of the papers listed in GC are duplicates.
3. There could also actually be several versions of the same paper available from different sources appearing as different papers.

Results

The h-index

Tables 2-4 show the main results for the three samples of academics. This includes: the years since first recorded publication as registered in GS (including published PhD thesis where appropriate); the largest number of citations for a publication; the h-index; the g-index; the contemporary h-index; and the h-rate. Table 5 shows summary statistics by group and overall

Tables 2-4 about here

Concentrating firstly on the h-index, we can see a range from 4 to 38. This is obviously quite large given that these are all senior academics, but not out of line with the studies of computer and information scientists discussed above (although they used WoS rather than GS and so will tend to have lower results). Given that this is a reasonably large sample of senior management academics we might hypothesise that these figures represent the maximum that is achievable in a career, and suggest that an h-index of over 20 (close to Q3 in the data) indicates a high level of research achievement. However, at the other end an h-index of below 10 (close to Q1) demonstrates relatively poor research productivity. To have, say, only 7 papers with 7 or more citations after a career of 20 or 30 years suggests that academics have been awarded Fellowships by learned societies not on the basis of their research but presumably because of their contribution in other ways. With a larger sample of academics, covering a wider range of career length, it may be possible to indicate what sort of h-index should be achieved by a particular career stage.

Table 5 about here

Looking at the groups, it can be seen that BAM and INFORMS have very similar means of 18 while COPIOR is less at 15. This is not unexpected since both BAM and INFORMS have larger scholarly communities than does COPIOR, BAM because of the range of disciplines and INFORMS because of the number of American academics. The range of values for COPIOR is actually quite good as it has the highest minimum, and second highest maximum, but it has a large density around the median of 11, while INFORMS has a significant density around 22 (Q3) as can be seen from Figure 1. The underlying reason for this is likely to be the greater population of US academics and, partly because of that, the tendency for US journals (in which UK academics seldom publish) to have much higher citation rates (Mingers and Harzing, 2007)

Figure 1 about here

We can also look at the relationship between the h-index and career-length as discussed by Burrell (2007b). As Burrell showed, the h-index is directly proportional to career length but what will differ between researchers is the rate of production of papers and citations, and hence the rate of growth of h. The final column of Tables 1-3 shows the h-rate (h-index/years) for each researcher. As can be seen, the maximum values for this dataset are just over 1 although the mean is around 0.6. So, a top researcher is adding to their h-index by about 1 per year. Less productive researchers have rates as low as 0.3.

Figure 2 about here

The results are graphed in Figure 2. Each point shows one of the sample members with their h-index against their career length. A straight line from a point to the origin would show their average h-rate. Those towards the x-axis have low rates while those towards the y-axis have the highest rates. In fact, the extreme points towards the y-axis can be regarded as forming a stochastic frontier, or as being output-efficient in DEA terms, as the line shows. In other words, given our sample the frontier shows the highest possible h-index (regarded as an output) for a given career length (input). The results show that for senior academics the frontier is at just over 1 per year. It would be useful to have a sample of junior academics to study the frontier in earlier years. The graph also shows that for any given career length there is a very wide variability of h-index, reflecting very different research productivity.

Table 6 about here

Table 6 shows the correlations between the main measures, career length, and maximum cites. All four measures are very highly correlated as would be expected. h is most highly correlated with hc, which is a simple variant of it. The correlation of h with Years is only moderate (0.33) reflecting the considerable scatter shown in Figure

2. The correlation with Max cites is also moderate (0.45) and this reflects the non-linearity of the relationship as theorised by Glänzel (2006a) who suggested a power relation. In fact regression estimates the relationship as:

$$\text{Max cites} = 2.05\ h^{1.6}$$

Both the hc-index (0.15) and the h-rate (-0.22) have low or negative correlations with Years since both emphasise more recent publications.

g-index

Whilst a strength of h is that it is not influenced by a few very highly cited works, some account should be taken of these and Egghe (2006) suggested the g-index instead of the h-index. We can see some relevant examples in the data: Peter Checkland has a reasonably high h-index of 25, but his first book, *Systems Thinking, Systems Practice* (Checkland, 1981) has received a massive 2700 citations. Should the extent of this contribution be ignored? Others in this category are: John Child, Peter Buckley, Frank Kelly, Kenneth Baker and Martin Puterman.

Clearly the g-index works in these cases. Checkland moves up from 6th to 2nd with a g of 83, significantly above the 3rd person (57). Child and Buckley move up a place but are already near the top. The biggest mover is Puterman who goes from 24th to 6th within INFORMS, and moves by 48 places when all three groups are combined. The mean change in rank in the combined group is 7.7.

Whether g is to be preferred to h is a matter of judgement. It does recognise classic, highly-cited works but as Puterman's case shows is advantageous to those with a single "big-hit", something the h-index was designed to avoid. g is also less intuitive than h and not as easy to see simply by inspection.

hc-index

The hc-index (Sidiropoulos et al., 2006) was designed to place more emphasis on current work as opposed to those who are no longer so active. One might therefore expect to see older academics falling down the rankings, and those at the most productive parts of their career moving up. This does happen but the changes are generally not as significant as with the g-index. Those moving up (in the combined group) include Susan Cartwright, Ken Starkey, Gerald Hodgkinson, and James Curran, all with short careers relative to this senior group. Those moving down include John Burgoyne, Michael Dempster, and Paul Williams. The mean change in rank was 7.6.

Again, this alteration to the basic h-index does work in changing the emphasis towards more current work, but also make the results more complex, and to some extent arbitrary depending on the parameters used. (This study used the same ones as in the original, discussed in section two above).

h-rate

Changing to the h-rate has the biggest effect on rankings with a mean change of 11.1. The effect is to improve those with short careers at the expense of those with long careers. Some examples in the combined group are: David Yao and David Simchi-Levy move to 1st and 2nd from 9th and 20th; Richard Whittington moves up 27 places; Sue Cox 40 and Richard Ormerod 45. Moving down are John Burgoyne (-21), Linus Schrage (-28) and Salah Elmaghraby (-33).

Using a rate does seem to have some positive features – it evens out the fact that the h-index always increases over time; there is evidence (Burrell, 2007b) that individual researcher's rate is roughly constant; and it would allow ECRs to compete on a more equal footing as their h-index may be increasing quickly even though it is still at a low absolute level. Intuitively, the rate of increase of h is as clear as the idea of the h-index itself.

h-index for research groups

One of the claimed strengths of the h-index is that it can be used at a variety of levels, e.g., research group, or journal. Considering using it to compare research groups, one would include all the publications for all members of the group and calculate h for that set of papers. However, in practice it would be strongly influenced by the group member with the highest h-index. This value of h would be increased to the extent that other group members had papers with more than h citations. Papers, and indeed members, who did not reach that threshold, would not contribute to the group-h at all. The effects of this are:

- The group-h would not reflect the size of the group or the total number of publications at all. This would seem undesirable as a measure of the overall strength of the group, especially as it would tend to exclude the contribution of early career researchers. There might be a group in which the papers of four long-serving members made up the group-h. The h value would not then distinguish whether the group also had several productive ECRs or no other members at all, yet in terms of sustainability the former is clearly to be preferred.
- If the group has one member who clearly dominates in terms of citations and publications then the group-h will really only be a reflection of that person's h. For instance, in a particular OR department one academic has an individual h of 28, the next nearest being 12. When the department is considered as a whole, the h value for the combined publications is 30. This is not much above the best individual one because there are few papers with more than 28 citations. Where it is more evenly spread the rise can be bigger, for example in a group where the best individual h's were 12, 10 and 7, the group-h was significantly higher at 15.

Van Raan (2005) studied the h-index for 147 chemistry groups, for which he had extensive and reliable publication data, and compared it with many traditional measures. He found that h was most strongly correlated with the total number of citations of the groups and was therefore best seen as a measure of the sheer outputs of the groups, being biased against small but highly productive ones. This does not necessarily tally with my comments above and more detailed research into this question would be useful.

One could overcome some of these effects by calculating the group-h as the mean of the individuals' h values but this would distort in the opposite direction towards small groups. My overall view is that h is not especially suitable for measuring the contribution of research groups.

h-index for journals

The h-index can also be used to assess the contribution of journals, again by considering the most highly cited of their papers perhaps over a set time period. This can be done easily in GS using *Publish or Perish*. There is considerable debate and contention about the subject of journal rankings (Mingers and Harzing, 2007). There are two main approaches: peer review where rankings are constructed by various academic communities, and revealed preference based on actual academic behaviour as revealed by metrics such as citations rates typically measured by the ISI impact factor (IF). Some studies combine the two (Baden-Fuller et al., 2000). Whilst impact factors are useful as a snapshot in time of the short-term impact of a journal's papers they are biased towards journals that publish up-to-the minute, often empirically-based reviews and against those which publish work of a more long-lasting nature the significance of which may take time to be recognised. Important papers are often cited for as long as 20 or 30 years and it has also been shown (Mingers, 2008) that there can be both "sleeping beauties" that get significant citations only several years after publication, and "shooting stars" which get many citations in the first few years and then disappear.

Some results for a selection of MS/OR journals are shown in Table 7. These are the h-index, the most highly cited paper, the 2006 impact factor and rankings from a statistical analysis combining peer review rankings with an impact factor (Mingers and Harzing, 2007).

Table 7 about here

We can see that the h-index gives an easily-understandable and robust quantitative measure that distinguishes clearly between the journals. The measure corresponds well with the journal ranking, the only real differences being *EJOR*. This has a high h, the most cited paper of all journals, and a good IF, and so arguably should be classified as a 4* journal although it has not been in any rankings I have seen. The other advantage of h over the ranking is that it reveals differences within a grade – for example *Management Science* is twice *Mathematical Programming* yet both are generally considered 4*. The impact factors vary from both h and the ranking reflecting perhaps their short-term nature. As an interesting comparison, within the field of neuroscience there are journals with massive impact factors and yet quite modest h-indices: for example, the *Annual Review of Neuroscience* has an IF of 28.5 but an h-index of only 146, and *Behavioral and Brain Sciences* has an IF of 15.0 but an h of 119 demonstrating presumably the speed with which knowledge develops in this area. Impact Factors can also vary significantly from year to year for the same journal..

Overall, the h-index seems to be a valuable addition to the many ways that journal quality is assessed.

Conclusions

The main conclusion of this study is that the h-index does provide a valuable and reliable measure of the research productivity of academics over a period of time. Its main advantages are that: it measures both quality and quantity in a single number; it is easily measured and also easily understood; it is very robust to the measurement problems that occur with publications and citations; and it correlates well with many other standard bibliometric measures. It also would appear to be a valid way of evaluating the quality of journals, but less so for assessing research groups.

There are limitations: it can only ever rise and so is biased towards long-standing researchers who can still improve their h-index even when they are no longer active; consequently, it does not measure well the contribution of early career researchers, who will automatically have a low h-index, or those who are currently very active. Modifications have been suggested to overcome these deficiencies and whilst these do work, they lose the simplicity and transparency of the standard h-index. Using the rate of h increase rather than the absolute value is an approach that may overcome these problems.

Overall, it may be that no one single measure can adequately reflect the full research performance of an individual or group and so it may be necessary to use a combination of measures, including the h-index. This could be approached through multi-attribute utility functions or possibly DEA.

More research is needed to:

- Study the h-index for larger and more diverse groups of researchers, especially those earlier in their career;
- Carry out more comparisons both across and within social science disciplines;
- Undertake comparisons with the more sophisticated bibliographic measures to validate the reliability of the h-index;
- Develop rigorous stochastic models to understand its theoretical properties.

Paper Number	W	X	Y	Z
1	19	33	54	43
2	14	28	43	13
3	8	8	13	8
4	8	3	8	7
5	8	2	7	5
6	5	2	2	
7	5	1	2	
8	3	1	2	
9	2	1	1	
10	2	1	0	
11	2	0		
12	2	0		
13	1	0		
14	1			
15	1			
16	0			
17	0			
Papers	17	13	10	5
Total cites	81	80	132	76
Mean cites	4.8	6.2	13.2	15.2
Max cites	19	33	54	43
h-index	5	3	5	5

Table 1 Illustration of Citation Measures for Four Academics

Name	Institution	Years	Max Cites	h-index	g-index	hc-index	h-rate
Cary Cooper	Lancaster Management School	38	217	38	59	22	1.00
Andrew Pettigrew	Bath School of Management	35	648	32	77	20	0.91
John Child	Birmingham Business School	40	1361	31	71	18	0.78
Peter J Buckley	Leeds University Business School	32	1058	26	57	15	0.81
John Storey	Open University Business School	31	448	26	48	16	0.84
Christopher Voss	London Business School	27	168	21	35	13	0.78
John Bessant	Imperial College London	29	851	20	45	15	0.69
John Burgoyne	Lancaster Management School	37	349	19	33	9	0.51
Nigel Nicholson	London Business School	32	166	19	32	11	0.59
David Buchanan	Cranfield School of Management	25	156	18	30	10	0.72
Mark Easterby-Smith	Lancaster Management School	27	655	18	44	12	0.67
Mike Wright	Nottingham Uni. Bus. School	22	475	18	37	15	0.82
Gordon Foxall	Cardiff Business School	35	120	17	28	9	0.49
David Otley	Lancaster Management School	32	211	17	36	11	0.53
Richard Whittington	University of Oxford	20	333	17	36	13	0.85
Andrew Kakabadse	Cranfield School of Management	30	62	16	23	12	0.53
John Saunders	Aston Business School	35	72	16	26	12	0.46
Howard Thomas	Warwick Business School	36	278	16	41	12	0.44
Susan Cartwright	Manchester Business School	18	124	15	30	12	0.83
Graham Hooley	Aston Business School	29	82	15	26	10	0.52
Gordon Greenley	Aston Business School	25	196	15	29	9	0.60
Ken Starkey	Nottingham Uni. Bus. School	20	138	15	31	12	0.75
James Curran	Goldsmiths, University of London	30	136	14	27	12	0.47
Gerard Hodgkinson	Leeds University Business School	20	69	13	22	11	0.65
Sue Cox	Lancaster Management School	16	16	12	20	10	0.75
Jean Hartley	Warwick Business School	27	129	11	22	9	0.41
Elizabeth Chell	University of Southampton	31	176	10	23	7	0.32
David Tranfield	Cranfield School of Management	35	121	10	18	7	0.29
Richard Thorpe	Leeds University Business School	25	655	8	30	7	0.32
Peter Mckiernan	University of St Andrews	27	56	6	13	4	0.22

Table 2 Citation Statistics for a Sample of BAM Fellows

Data collected from Google Scholar during July and August 2007

Name	Institution	Years	Max Cites	h-index	g-index	hc-index	h-rate
Frank Kelly	University of Cambridge	32	1430	34	88	21	1.06
John Beasley	Brunel University	31	466	29	57	18	0.94
John Mingers	University of Kent	27	252	28	48	16	1.04
Colin Eden	University of Strathclyde	32	226	28	50	16	0.875
Chris Potts	Southampton	29	171	26	45	13	0.90
Peter Checkland	Lancaster University	38	2692	25	83	12	0.66
Michael Dempster	University of Cambridge	39	87	19	29	9	0.49
Emmanuel Thanassoulis	Aston University	24	159	18	35	10	0.75
Derek Bunn	London Business School	32	102	18	31	12	0.5625
Robert Dyson	University of Warwick	34	159	18	33	11	0.52941
Jonathan Rosenhead	LSE	39	298	15	30	8	0.38
Mike Pidd	Lancaster University	30	372	15	31	10	0.50
Robert Fildes	Lancaster University	31	203	15	26	9	0.48
Paul Williams	LSE	33	484	13	32	6	0.39
Val Belton	University of Strathclyde	26	216	13	29	9	0.50
Said Salhi	University of Kent	20	35	12	17	8	0.60
Gautam Mitra	Brunel University	39	102	11	19	7	0.28
Nigel Meade	Imperial College London	29	77	11	19	7	0.38
Gwyn Bevan	LSE	31	29	11	17	8	0.35
Kevin Glazebrook	Lancaster University	31	94	10	17	8	0.32
Richard Eglese	Lancaster University	21	205	10	21	7	0.48
Richard Ormerod	University of Warwick	15	45	10	15	7	0.67
KS Hindi	Brunel University	29	26	10	14	6	0.34483
Vitaly Strusevich	University of Greenwich	18	54	10	16	6	0.55556
AP Muhlemann	University of Bradford	35	54	10	16	6	0.28571
Cecilo Mar-Molinero	University of Kent	28	25	9	14	5	0.32
Tim Bedford	University of Strathclyde	22	157	8	17	7	0.36364
KAH Kobbacy	University of Salford	18	23	7	11	4	0.39
BC Dangerfield	University of Salford	18	22	7	10	4	0.38889
K Darby-Dowman	Brunel University	27	32	7	11	3	0.25926

Table 3 Citation Statistics for a Sample of COPIOR Members

Data collected from Google Scholar during July and August 2007

Name	Institution	Years	Max Cites	h-index	g-index	hc-index	h-rate
BL Golden	University of Maryland	33	209	28	51	14	0.85
Paul Zipkin	Duke University	30	368	28	49	14	0.93
David D. Yao	Columbia University	25	131	28	40	18	1.12
Kenneth R.Baker	Dartmouth College	40	1115	26	59	17	0.65
Dorit S.Hochbaum	University of California	27	874	26	55	14	0.96
Michael J.Todd	Cornell University	34	218	25	51	18	0.74
John M.Mulvey	Princeton University	32	247	25	42	13	0.78
Michael Florian	University of Montréal	38	153	25	39	11	0.66
Linus E.Schrage	University of Chicago	46	890	23	53	12	0.50
David Simchi-Levi	MIT	21	381	23	50	17	1.10
Robert J.Vanderbei	Princeton University	27	316	23	49	16	0.85
David F.Shanno	Rutgers University	37	234	22	46	13	0.59
Shmuel S.Oren	U. California at Berkeley	35	80	22	33	13	0.63
Mark S Daskin	Northwestern University	29	293	19	37	12	0.66
Bruce W.Schmeiser	Purdue University	32	164	18	32	10	0.56
Leon S.Lasdon	University of Texas	43	510	17	42	8	0.40
Salah Elmaghraby	North Carolina State	49	205	17	30	8	0.35
DW Hearn	University of Florida	35	74	16	28	11	0.46
Leroy B.Schwarz	Purdue University	36	110	15	27	8	0.42
Andres Weintraub	Universidad de Chile	31	43	15	20	10	0.48
Mordecai Avriel	Israel Institute of Technology	41	489	14	35	8	0.34
Robert D. Doverspike	AT&T Labs Research	19	107	14	25	10	0.74
Martin L.Puterman	Uni. of British Columbia	35	1261	13	50	8	0.37
Gerald G Brown	Naval Postgraduate School	36	329	13	29	8	0.36
Chelsea C.White III	Georgia Inst. of Technology	31	38	12	18	7	0.39
James E. Matheson	Stanford University	39	458	10	30	7	0.26
L Robin Keller	University of California	25	35	10	15	6	0.40
Peter C. Bell	Uni. of Western Ontario	30	63	9	15	7	0.30
Kalyan Singhal	University of Baltimore	29	27	6	10	5	0.21
Vicki L.Sauter	Uni. of Missouri-St.Louis	27	32	4	11	3	0.15

Table 4 Citation Statistics for a Sample of INFORMS Fellows

Data collected from Google Scholar during July and August 2007

Variable	Group	Mean	StDev	Minimum	Q1	Median	Q3	Maximum
h_index	1	18.07	7.19	6.00	13.75	17.00	20.25	38.00
	2	14.60	7.22	7.00	10.00	11.50	18.00	34.00
	3	18.20	6.88	4.00	13.00	17.50	25.00	28.00
	All	16.96	7.22	4.00	11.00	16.00	22.00	38.00
g_index	1	36.00	15.39	13.00	25.25	31.50	44.25	77.00
	2	28.13	19.43	10.00	15.75	20.00	32.25	88.00
	3	35.70	14.06	10.00	26.50	36.00	49.25	59.00
	All	33.28	16.67	10.00	19.75	30.00	44.25	88.00
hc_index	1	12.23	4.10	4.00	9.00	12.00	15.00	22.00
	2	8.80	4.06	3.00	6.00	8.00	10.25	21.00
	3	10.86	4.00	3.00	8.00	10.50	14.00	18.00
	All	10.63	4.25	3.00	7.00	10.00	13.00	22.00

1 = BAM, 2 = COPIOR, 3 = INFORMS

Table 5 Summary Statistics for the Three Samples

	Years	Max Cites	h-index	g-index	hc-index	h-rate
Years	1					
Max Cites	0.320469	1				
h-index	0.331489	0.454057	1			
g-index	0.372514	0.766001	0.88685	1		
hc-index	0.154045	0.396634	0.921536	0.834678	1	
h-rate	-0.21614	0.272067	0.830991	0.69369	0.862511	1

Table 6 Correlations between Measures across all Groups (n=90)

Journal	h-index*	Max Cites*	2006 Impact Factor	Kent Ranking**
Management Science	169	2050	1.687	4*
Operations Research	127	917	0.615	4*
European J. Operational Research	98	3409	0.918	3*
Mathematical Programming	84	1003	1.117	4*
Decision Sciences	63	423	1.620	3*
Annals of OR	63	402	0.589	3*
J. Optimization Theory and Applications	58	534	0.633	3*
J. Operational Research Society	58	1508	0.597	2*
Computers and OR	51	248	0.893	2*
Omega	46	326	0.663	2*

* over the length of publications available to Google Scholar

** (Mingers and Harzing, 2007) available at http://www.kent.ac.uk/kbs/staff_detail.php?page_id=15&ID=83

Table 7 h-index for Journals

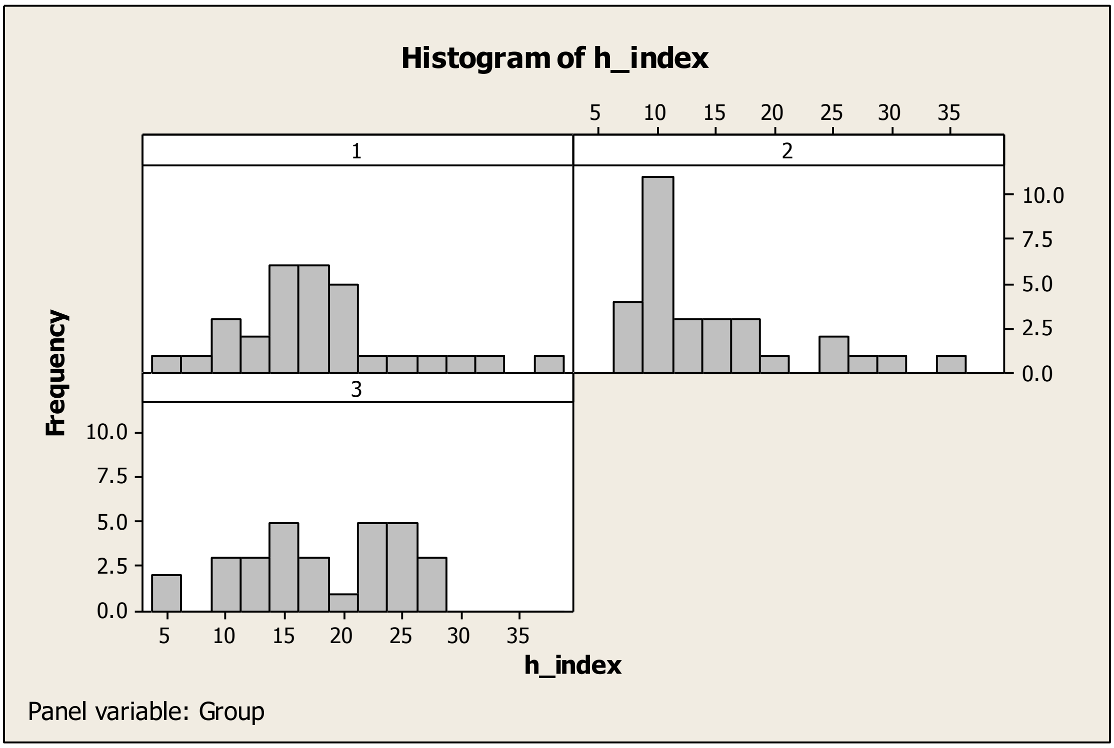

Figure 1 Histograms of the h-index for BAM (1), COPIOR (2) and INFORMS (3)

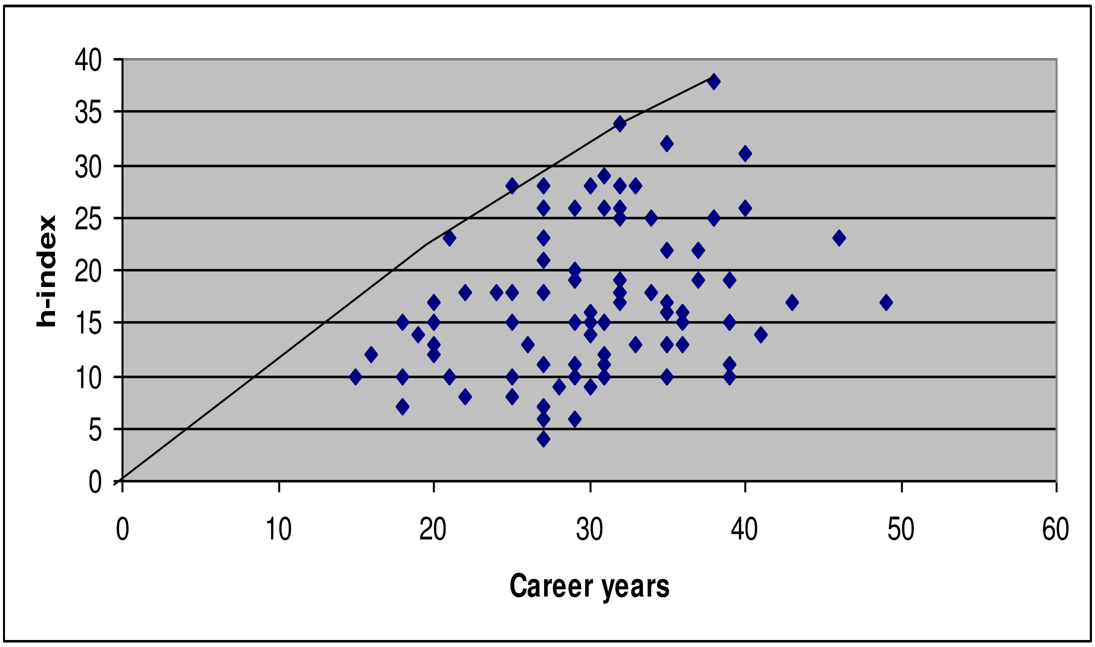

Figure 2 h-index Plotted against Career Years for All Researchers showing the Stochastic Frontier

References

Baden-Fuller, C., Ravazzolo, F. and Schweizer, T. (2000). Making and measuring reputations - The research ranking of European business schools. Long Range Planning **33**: 621-650.

Batista, P., Campiteli, M., Kinouchi, O. and Martinez, A. (2006). Is it possible to compare researchers with different scientific interests? Scientometrics **68**: 179-189.

Bornman, L. and Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work? Scientometrics **65**: 391-392.

Brumback, R. (2008). Worshipping false idols: the impact factor dilemma. J. Child Neurology **23**: 365-367.

Burrell, Q. (2007a). Hirsch's h-index: a stochastic model. J. of Informetrics **1**: 16-25.

Burrell, Q. (2007b). Hirsch index or Hirsch rate? Some thoughts arising from Liang's data. Scientometrics **73**: 19-28.

Checkland, P. (1981). Systems Thinking, Systems Practice. Wiley: Chichester.

Cronin, B. and Meho, L. (2006). Using the h-index to rank influential information scientists. J. Am. Soc. for Information Science and Technology **57**: 1275-1278.

DfES (2006). Reform of Higher Education Research Assessment and Funding. DRN05/13, Department for Education and Skills, London.

Egghe, L. (2006). Theory and practice of the g-index. Scientometrics **69**: 131-152.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. Science **178**: 471-479.

Glänzel, W. (2006a). On the h-index - a mathematical approach to a new measure of publication activity and citation impact. Scientometrics **67**: 315-321.

Glänzel, W. (2006b). On the opportunities and limitations of the h-index. Science Focus **1**: 10-11.

Glänzel, W. and Schubert, A. (1988). Theoretical and empirical studies of the tail of scientometric distributions. In Egghe, L. and Rousseau, R. (Eds.), Informetrics 87/88. Elsevier, New York, pp. 75-83.

Harzing, A.-W. (2007a) Publish or Perish, http://www.harzing.com/pop.htm Access: June 2007.

Harzing, A.-W. (2007b). Reflections on Google Scholar, http://www.harzing.com/pop_gs.htm, Access: March, 2007.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America **102**: 16569-16572.

Iglesias, J. and Pecharroman, C. (2006). Scaling the h-index for different scientific ISI fields, arXiv:physics/0607224v1 [physics.soc-ph], Access: September, 2007.

Jennings, C. (1998). Citation data: the wrong impact? Nature Neuroscience **1**: 641-642.

Kelly, C. and Jennions, D. (2006). The h index and career assessment by numbers. TRENDS in Ecology and Evolution **21**: 167-170.

Liang, L. (2006). h-index sequence and h-index matrix: constructions and applications. Scientometrics **69**: 153-159.

Macdonald, S. and Kam, J. (2007). Ring a ring o'roses: quality journals and gamesmanship in Management Studies. J. Management Studies **44**: 640-655.

MacRoberts, M. and MacRoberts, B. (1987). Problems of citation analysis: A critical review. J. Am. Soc. for Information Science **40**: 342-349.

Meho, L. and Yang, K. (2007). A new era in citation and bibliometric analyses: Web of Science, Scopus and Google Scholar. J. American Society for Information Science and Technology **58**: 2105-2125.

Mingers, J. (2008). Exploring the dynamics of journal citations: modelling with S-curves. J. Operational Research Society **e-publication**, 10.1057/palgrave.jors.2602428.

Mingers, J. and Harzing, A.-W. (2007). Ranking journals in business and management: a statistical analysis of the Harzing dataset. European J. Information Systems **16**: 303-316.

Oppenheim, C. (2007). Using the h-index to rank influential British researchers in Information Science and Librarianship. J. Am. Soc. for Information Science and Technology **58**: 297-301.

Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. Scientometrics **69**: 117-120.

Seglen, P. (1997). Why the impact factor of journals should not be used for evaluating research. British Medical Journal **314**: 498-502.

Sidiropoulos, A., Katsaros, D. and Manolopoulos, Y. (2006). Generalized h-index for disclosing latent facts in citation networks, arXiv:cs.DL/0607066 v1, Access: May, 2007.

van Raan, A. (2005). Comparison of the hirsch-index with standard bibliometric indicators and with peer judgement for 147 chemistry research groups. Scientometrics **67**: 491-502.

van Raan, A. J. (2003). The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. Technology Assessment - Theory and Practice **1**: 20-29.

Walters, W. (2007). Google Scholar coverage of a multidisciplinary field. Information Processing and Management **43**: 1121-1132.