Exploring the Dynamics of Journal Citations: Modelling with s-Curves

Author(s): J. Mingers

Source: *The Journal of the Operational Research Society*, Aug., 2008, Vol. 59, No. 8 (Aug., 2008), pp. 1013-1025

Published by: Palgrave Macmillan Journals on behalf of the Operational Research Society

Stable URL: https://www.jstor.org/stable/20202165

Journal of the Operational Research Society (2008) 59, 1013–1025 0160-5682/08 $30.00

www.palgrave-journals.com/jors

Exploring the dynamics of journal citations: Modelling with s-curves

J Mingers*

University of Kent, Canterbury, Kent, UK

This paper reports on an exploratory analysis of the behaviour of citations for management science papers over a 14-year period. Citations often display s-curve type behaviour: beginning slowly, rising in response to previous citations, and then declining as the material becomes obsolete. Within the context of citation research such functions are known as obsolescence functions. The paper addresses three specific questions: (i) can collections of papers from the same journal all be modelled using the same obsolescence function? (ii) Can we identify specific patterns of behaviour such as 'sleeping beauties' or 'shooting stars'? (iii) Can we predict the number of future citations from the pattern of behaviour in the first few years? Over 600 papers published in six leading management science journals are analysed using a variety of s-curves.

Journal of the Operational Research Society (2008) **59,** 1013–1025. doi:10.1057/palgrave.jors.2602428
Published online 30 May 2007

Keywords: journal citations; management science journals; RAE; s-curve

1. Introduction

There is currently much interest in measuring the quality of academic research whether at the institutional, journal, or personal level. In the main, this has been done through peer review (also known as stated preference) where a group of academics produce a ranking of journals. Many of these rankings have been collected together on a website by Harzing (2005) and a combined ranking based on statistical analysis has been produced by Mingers and Harzing (2005). The alternative approach is to use revealed preference measures based on actual publication behaviour, especially using the paper citation data available from the ISI. Interest is particularly high at this point in time in the UK where it is being proposed that the long-standing Research Assessment Exercise (RAE) should be replaced by a process based on metrics (DfES, 2006).

Broadly, citation analysis began in 1961 with the publication of the *Science Citation Index* (*SCI*) and has grown significantly since, although there have always been criticisms of the reliability of measures based on ISI citations. Most studies use the total number of citations received by papers or journals to judge their quality, or they may use the Garfield impact factor published by ISI (Garfield, 1972), which measures the number of citations per paper received in the last 2 years. While they can be useful many people have pointed out the potential biases in such measures, especially the type of article, coverage of the topic, language, and disciplinary area (MacRoberts and MacRoberts, 1987; Seglen, 1997; Jennings, 1998; Callahan *et al*, 2002; Glänzel and Moed, 2002).

Within the literature on citations we can see three streams of work: papers using numbers of citations to rank journals or departments; papers investigating factors that affect the numbers of citations; and papers directly analysing the generation process and dynamic behaviour of citations.

In the Business and Management field, Doyle and Arthurs (1995) used 10-year citation rates to determine the most influential journals (largely US) and then ranked business schools in terms of their publication in these journals. This paper triggered a heated debate about the relative merits of citation studies against peer review (Doyle *et al*, 1996a,b; Jones *et al*, 1996a,b). Baden-Fuller *et al* (2000) identified 32 top journals through citations as part of a review of the rankings of business schools, which was combined with peer review. Management journals have also been ranked using citations—by Tahai and Meyer (1999) who ranked 65 top journals based on citations from 17 key management journals, and DuBois and Reeb (2000) who combined citations and a survey to rank international business journals.

In terms of the causal influences on the number of citations, Bettancourt and Houston (2001) investigated factors affecting the number of citations in marketing journals, in particular the type of article (e.g, mathematical theory, verbal theory, empirical, or methodological) and the subject area. van Dalen and Henkens (2001) studied a wide range of possible explanatory variables grouped in terms of characteristics of the journal, of the article itself, and of the authors. They found that the major effect was the perceived quality of the journal.

The current paper contributes to the third stream, an analysis of the behaviour of citations in themselves. If we consider

*Correspondence: *J Mingers, Kent Business School, University of Kent, Canterbury CT2 7NZ Kent, UK.*
E-mail: j.mingers@kent.ac.uk

a single paper we can trace the number of citations over time. We can also do this for a cohort of similar papers, e.g from the same journal, subject or time period. Several models for the process generating citations have been proposed. Glänzel and Schoepflin (1995) used a linear birth process while Egghe (2000) assumed that the citation process was exponential. The most common and intuitive view is to see the citations as essentially random with some average rate and thus model their generation as a Poisson process. If we move from a single paper to a cohort or collection then each paper will have a different mean rate and this leads to a non-homogeneous Poisson process (NHPP). Burrell (1990, 2001) assumed a gamma distribution of rates across the collection leading to the gamma-Poisson process, which results in a negative binomial distribution of citations per paper from a particular source.

However, the mean citation rate varies over time as well as across papers. Typically, although there is a degree of variation, the number of citations for a paper is small to begin with, rises to a peak in response to other citations, and then subsides as the paper's material becomes obsolete (Cunningham and Bocock, 1995; Redner, 2005). This is very similar to the s-shape curve found in new product diffusion and technological forecasting. Burrell (2003) incorporated the general concept of an obsolescence function into the gamma-Poisson model but did not specify a particular functional form. Mingers and Burrell (2006) successfully fitted this gamma-Poisson model incorporating obsolescence to a sample of over 600 papers from the management science literature and were able to draw some tentative conclusions about possible obsolescence functions. The details of this model are reproduced in Appendix A. The work that this paper describes follows on with direct empirical testing of various forms for the obsolescence function on the same sample of data.

In particular, the paper is motivated by three questions:

1. To what extent can collections of papers (e.g, all from one journal) be modelled by the same obsolescence function? If so, what are the most suitable functions? Our hypothesis would be that we require a range of functions to fit different citation patterns.
2. Can we identify different patterns of behaviour from the standard s-curve? For example, 'sleeping beauties' (Van Raan, 2004) which remain uncited for some time before suddenly becoming popular, perhaps because they were ahead of their time; or 'shooting stars' which are heavily cited initially but die quickly perhaps because they were part of a fad. To what extent do we find other anomalies, for example papers that are never cited (Rousseau, 1994) or seminal papers that never become obsolete (Redner, 2005)? If new patterns emerge can we explain what generates them?
3. To what extent can the number of future citations be predicted given the pattern of citations over the first few years? This is of particular interest for quality exercises such as the UK's RAE where all the papers evaluated will be less than 7 years old and so will still be young within their citation lifespan.

To address these questions a study was undertaken to fit a variety of s-curves to the citation histories of management science papers. The next section discusses the approach taken, and the third section reports on empirical results.

2. Methodology

The data set consisted of a sample of over 600 papers published in 1990 giving a 14-year history of citation behaviour. When the sample was taken this was felt to be long enough for most papers to have completed their citation lives but not so long that there would have been significant changes in academic citation behaviour. However, once the data were analysed it became clear that this was actually too short a period. The sample was all those papers published in six leading management science journals: *Management Science* (ManSci), *Operations Research* (OpsRes), *Decision Science* (DecSci), *European Journal of Operational Research* (EJOR), *Journal of the Operational Research Society* (JORS), and *Omega* (Omega). These were selected for their variety on several factors—level of prestige and quality; prevalence of heavily mathematical articles; US *versus* European; narrowness and width of coverage. The number of citations for each journal over the full 14-year period 1991–2004 was tabulated and the means and standard deviations are shown in Table 1, and histograms in Figure 1.

Two comments should be made about the data. (i) At first all document types from a journal were recorded. However, with JORS and EJOR there were large numbers of book reviews that virtually all received zero citations. While a book review could be cited it is very rare. Other journals, especially ManSci and OpsRes did not have book reviews and so had a much smaller proportion of zero citations. To avoid this bias, only documents of type article, editorial or letter were recorded. (ii) With the ISI database selection of a year in the database limits does not correspond exactly with the actual year of publication. For JORS for example, selecting '1990' picks up some papers from the end of 1989 and excludes some from the end of 1990. This required considerable manual intervention.

The mean number of citations (over 14 years) varied significantly from seven (JORS and Omega) to 39 (ManSci). All the distributions were extremely skewed with variances up to

Table 1 Summary statistics for the number of citations per paper

	Omega	*JORS*	*DecSci*	*EJOR*	*OpsRs*	*ManSci*
Number	51	123	43	202	112	85
Mean	7.2	7.3	11.1	11.3	14.7	38.6
Std. Dev.	15.5	17.9	14.0	19.0	28.6	42.4
Skewness	4.2	7.4	2.1	4.2	7.3	1.9
% zero cites	22	18	12	14	10	5
Max. cites	87	176	66	140	277	181

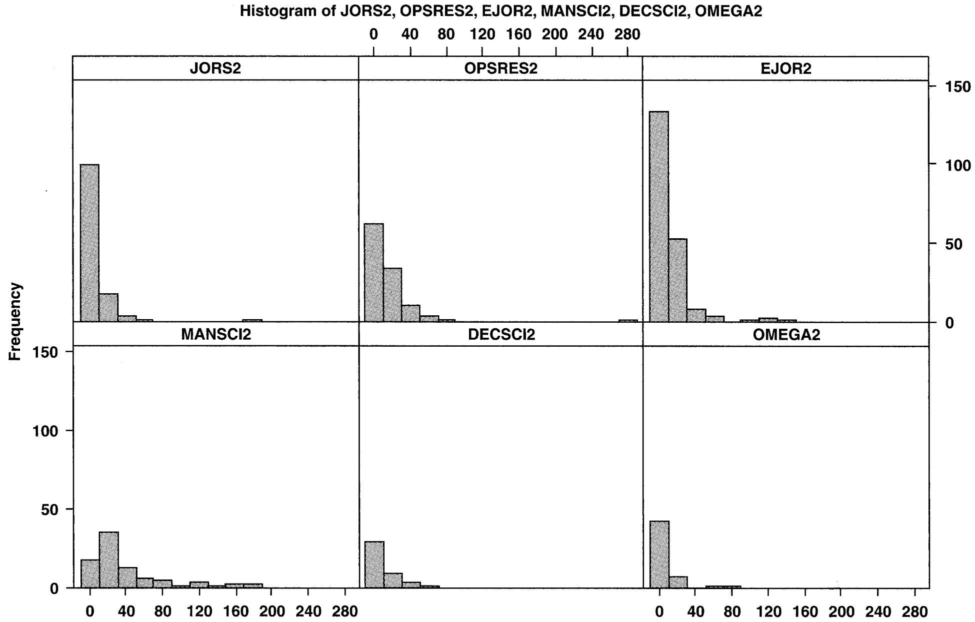

Figure 1 Histograms of the distributions of number of citations.

40 times the mean. The coefficients of skewness are all large given that values over 1 represent highly skewed distributions. The maximum number of citations for a paper ranged from 66 (DecSci) to 277 (OpsRes) although these were to some extent outliers. One interesting, and perhaps surprising, fact is the number of papers that were *never* cited during the period of observation. In each case (except ManSci), the modal value of the distribution of number of citations was in fact zero and the % of zero cites ranges from 5% (ManSci) to 18% (JORS) and 22% (Omega).

The next stage was to decide which s-curves were to be fitted to the data. Within the domains of technological forecasting and marketing, where s-curve fitting has primarily developed, the most common curves used are the logistic and the Gompertz (Martino, 1983). However, there are many curves that could be used and Meade and Islam (1998) identify 29 different ones. Indeed, almost all probability distributions have appropriately shaped cumulative curves and therefore could be used as an s-curve. Where such curves have been used, it is simply because they have an appropriate shape not because they are also probability distributions.

Meade and Islam classify their models into three classes—symmetric, asymmetric and flexible depending on behaviour around the point of inflection. The point of inflection of the s-curve (which is a cumulative curve) is equivalent to the point of maximum citation generation—that is, the mode of the corresponding probability density function. Symmetrical models have a fixed point of inflection which occurs at 50% of the eventual total citations. The growth and decline are symmetrical about this point. Asymmetrical models typically have their inflection point at less than 50% with a faster growth than decline. The underlying pdf is positively skewed. Flexible models can have variable inflection points, some even being greater than 50%. The underlying pdf's can take on a range of shapes both symmetrical and skewed. For this research one member of each of these classes was selected—the Pearl logistic (symmetrical), the Gompertz (asymmetrical) and the Weibull (flexible). The gamma distribution was also included as this had been used in the previous work (Mingers and Burrell, 2006). The gamma is also a flexible distribution similar in many respects to the Weibull. Details of these curves are shown in Appendix B.

The fitting process was straightforward—parameter values were estimated for each curve that minimized the sum of squared errors (SS) from the empirical cumulative citation data. All the curves require three parameters—one governing the shape, one the scale, and one the upper limit. This method is biased towards the higher values of the curve but this was felt to be acceptable especially given the importance and difficulty of estimating the upper limit of the s-curves (see below). There was a practical problem in that sometimes the solver would become stuck in a local optimum for a particular curve. This was easily detectable as the fitting was done manually and graphs showed where the curves were not fitted properly. Manually setting starting values always resolved the problem. In many cases, the fitting process was duplicated and the results were within 0.01%. There was no need to use more sophisticated methods such as AIC (Akaike, 1974) or BIC (Schwarz, 1978) since we were comparing models that all had the same number of parameters on the same data sets.

A further validation method was used by comparing the fitted parameters for different functions. The Weibull and the gamma, particularly, tend to give very similar results. In each case, it is possible to estimate the time of inflection from the fitted parameters (see Appendix B for the formula) and a plot of one against the other forms an almost perfect straight line ($r = 0.996$). Any deviations were investigated and re-fitted. A similar approach was used with the Pearl and Gompertz although the relationship had more variability.

3. Empirical results

3.1. *Fitting s-curves to collections of papers*

One of the first steps was to simply look at the pattern of citations over time for each journal and the total. This is shown in Table 2.

A bar chart is shown in Figure 2.

All the journals show a clear pattern of citations as expected. Citations rise to a peak most commonly after about 6 years, although there is some variation between journals. There is then something of a plateau before numbers begin to reduce. ManSci, particularly, maintains a high level for a long period—from 262 in period 4 to 265 in period 14. It was certainly surprising to me that citations should still be high over such a long period. Indeed, if one projects forward (see later) 15 years there would still be around 120 citations in total.

These data were the first to be used for s-curve fitting, initially for the total set of citations, and then splitting it down by journal, to begin to answer the first of the research questions.

Table 3 shows, for each journal separately and for the total citations, the fitted parameters for the four different curves. It also shows the sum of squares (ss) as a measure of goodness of fit.

Looking firstly at the total citations column, the fitted curves are shown in Figure 3. It appears from this that all the curves fit reasonably well although they give quite different future projections (the 'limit' value) with the Pearl logistic the lowest and the gamma the highest. However, looking in more detail at the year by year citations in Figure 4 shows that certain curves fit the data much better than others. The

Table 2 Number of citations in each year after publication (shaded boxes show modal period) for all papers, by journal

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
EJOR	55	74	97	161	183	253	221	175	208	156	151	164	133	157
DecSci	13	24	27	48	32	45	43	41	25	32	38	34	26	29
Jors	20	48	58	55	70	71	85	68	85	50	69	66	50	69
Mansci	60	182	186	262	249	256	252	241	253	217	275	232	232	265
Omega	6	23	21	26	29	43	36	19	34	29	33	30	17	11
OpsRes	51	84	91	126	137	167	126	126	113	107	107	113	93	104
Total	205	435	480	678	700	835	763	670	718	591	673	639	551	635
Cum. Total	205	640	1120	1798	2498	3333	4096	4766	5484	6075	6748	7387	7938	8573

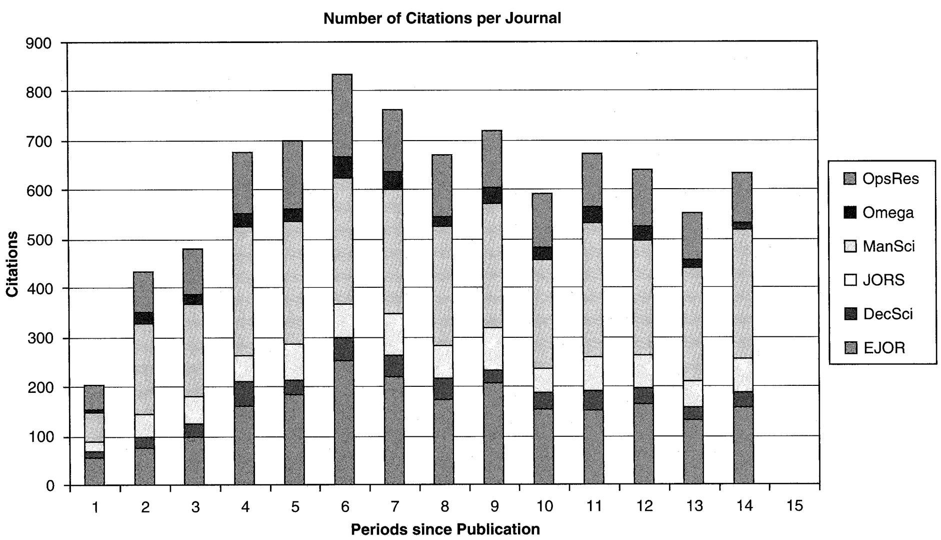

Figure 2 Number of citations in each year after publication for all papers, by journal.

Table 3 S-curves fitted to the collections of journals (shaded boxes show the best fits)

Curve	*Params.*	*OpsRes*	*Omega*	*ManSci*	*JORS*	*DecSci*	*EJOR*	*Total*
Pearl-logistic	scale	18.09	21.43	18.64	19.96	18.19	25.94	20.16
	shape	0.40	0.41	0.36	0.39	0.40	0.43	0.39
	limit	1578.80	376.14	3383.05	903.33	469.41	2226.71	8917.53
	ss	31314.70	1468.18	145425.72	7844.39	3057.25	47767.02	880015.91
Gompertz	scale	3.92	4.24	3.89	4.03	3.94	4.67	4.08
	shape	0.21	0.22	0.18	0.20	0.21	0.23	0.20
	limit	1831.38	437.01	4147.81	1085.36	544.41	2596.53	10600.43
	ss	7331.01	436.31	45501.04	1778.79	836.72	8654.29	191803.99
Weibul	scale	11.42	10.51	16.95	12.87	11.51	10.20	12.30
	shape	1.55	1.72	1.45	1.57	1.55	1.87	1.60
	limit	2048.26	449.88	5896.19	1258.52	611.16	2579.88	11998.79
	ss	2250.79	290.58	6349.23	470.83	283.78	7249.18	39264.97
Gamma	scale	7.00	0.58	12.89	8.41	7.03	4.50	7.53
	shape	1.78	2.05	1.57	1.76	1.78	2.39	1.83
	limit	2323.76	517.82	7157.49	1503.47	692.85	2931.90	14040.39
	ss	1879.95	320.53	4951.50	479.20	237.28	6003.11	28648.70

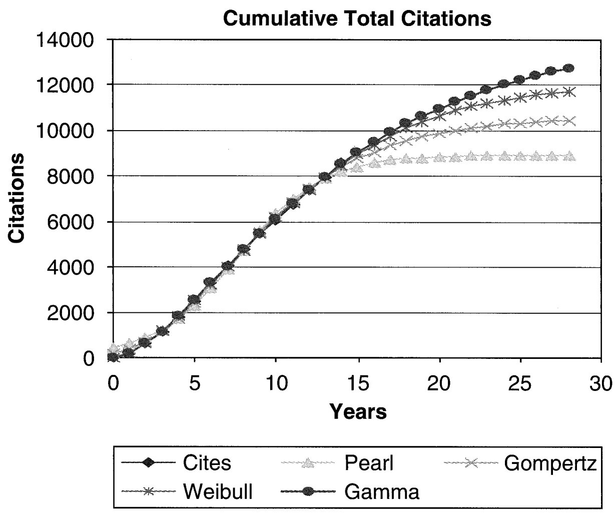

Figure 3 Fitting s-curves to the cumulative citations for all journals.

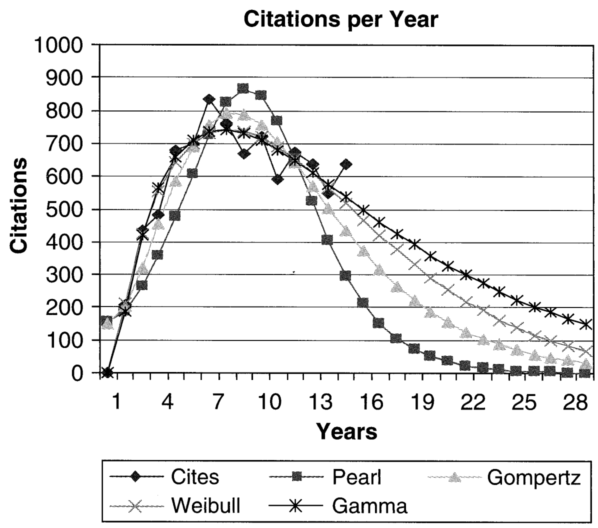

Figure 4 Citations per year for all journals.

Pearl curve is always a symmetrical curve with the point of inflection (maximum growth) occurring when it is at 50% of its maximum value (see Appendix A). Clearly the citation data are not symmetrical but strongly positively skewed. The Gompertz curve is not symmetrical but it also has a fixed inflection point at the same time as the Pearl curve but with a lower cumulative value. This too does not fit the data well.

In contrast, both the Weibull and gamma curves are very flexible in their shapes and points of inflection and can be fitted well to the data. This is reflected in the SS where the Pearl is worst with 880k, then the Gompertz with 192k, the Weibull 39k, and the gamma is best with 29k. The gamma estimate of the eventual total is 14 000 (after about 50 years) while the Weibull is 12 000 (after 40).

Moving down to the individual journals, we can see (in Table 3) that in four out of the six the gamma is best, with the Weibull being marginally best in the other two—Omega and JORS.

These results are broadly in line with those of Mingers and Burrell (2006). In that paper, obsolescence functions (ie, s-curves) were not fitted directly to the data but instead to curves derived from the additional citations distribution tested in the paper. This in turn was based on a gamma-Poisson model for the underlying process of citation generation combined with some form of obsolescence function, the particular form being estimated from the data. Several different obsolescence functions were tested including those used in this paper. The results showed the gamma function to

Table 4 Proportion of papers still active or with citations complete

	ManSci (%)	*JORS (%)*	*EJOR (%)*	*OpsRes (%)*	*DecSci (%)*	*Omega (%)*	*Total*
< 15	37	91	81	75	77	90	469
Still active	55	5	14	16	18	6	110
Complete	8	4	5	9	5	4	36

be the best overall in terms of lowest total SS, and individually best for ManSci. The Weibull was the second best overall and individually best for OpsRes and EJOR. The general conclusion, therefore, is that the gamma distribution is a good fit for collections of papers that are likely to show significant skew over time. The other interesting result is simply the length of time over which well-cited papers carry on getting cited.

3.2. *Fitting s-curves to individual papers*

The next stage of the analysis was to move down a level to individual papers and see whether particular patterns emerged at this level in order to address research questions one and two.

The first problem was which papers to analyse. Clearly there was no point in using papers that were very rarely cited and as a cut-off only those with 15 or more citations were considered—equivalent to averaging one per year. Although this seems fairly modest it removed over 75% of the papers as shown in Table 4. This differed significantly between journals—37% for ManSci, but over 90% for JORS and Omega.

On fitting curves to the remaining papers a further problem emerged—estimates of the upper bound. In some series, especially where citations rates were still high and it was not clear from the data whether the turning point had yet been reached, the different curves would generate hugely different upper limits. Sometimes these would be four or five times the number of citations so far recorded. This is generally a significant problem in fitting s-curves to sets of data that are not yet complete. In fact, Martino (1983), within the context of technological forecasting, argues that the upper limits should always be determined manually having regard to ultimate technological or economic constraints rather than be estimated from the data.

However, in our case this does not seem possible. How could we put a sensible limit on the total number of citations that any paper could possibly receive? While most highly cited management papers would go into the hundreds (papers in some areas of natural science go very much higher (Redner, 2005)), a particular paper could go into the thousands. Equally, it seems impossible to decide on some arbitrary length of time after which citations would finish. Experiments were tried constraining the estimate of the upper limit to two or three times the current level but these just distorted the fitting and still seemed essentially arbitrary. It was therefore decided at this stage to limit our analysis to those papers whose citation history appeared complete—that is, which were getting almost no further citations by year 15. The specific criterion used was that a paper would be considered still active if it had more than one citation in the last 2 years or more than two citations in the last 3 years (as well as more than 15 citations overall).

As can be seen from Table 4, this left only 36 papers in total with 15 or more citations considered to be completed. These were fitted to all four s-curves by minimizing the sum of errors across all 14 points. In looking at the results, we should bear in mind the following:

- The curves are all characterized by three parameters. The limit relates to the total number of citations—that is, the Y-axis. The scale parameter relates to time—the larger the value the longer the period over which citations occur. The shape parameter relates to spread and skewness or symmetry. The Pearl logistic can only be symmetrical; the Gompertz and gamma can be symmetrical or have positive skew; and the Weibull can also have negative skew.
- This sample of papers is obviously not representative of the whole as they have all been completed in a relatively short space of time. Those that are still active will generally be more skewed as their publications carry on.

After fitting, several different patterns of citations could be seen, although it should be emphasized that the data generally included a considerable degree of randomness. To illustrate the degree of variety of the curves that were found to fit, Figure 5 shows a plot of the scale and shape parameters for the Weibull distribution as fitted to the 36 different data sets. The points are also marked according to which function fitted best in terms of the sum of squares. For example, the two points in the extreme south-east corner had Weibull parameters of 0.8 and 38, and were best fitted by the gamma distribution.

The shape parameter governs the shape of the Weibull pdf (and thus also the shape of the cumulative curve that we are using). A value of 1 is equivalent to an exponential distribution which has a mode at 0 and strong positive skew. As the shape parameter increases (say between 2 and 4), the distribution becomes more symmetrical, while for higher values the distribution becomes very 'pointed' and even has negative skew. So Figure 5 shows that the shapes ranged from those in the S-E corner, already mentioned, which are highly (positively) skewed (shape < 1) and are cited over a long period; the majority which are relatively symmetrical and with shorter timescales, to those in the N-W corner which may

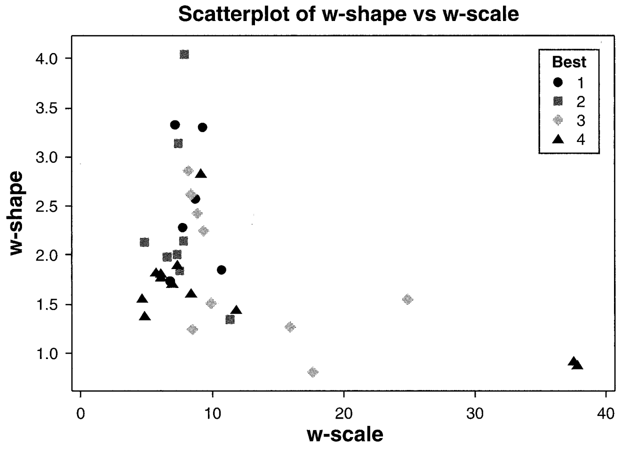

Figure 5 Fitted Weibull parameters for the completed papers 1 = Pearl, 2 = Gompertz, 3 = Weibull, 4 = gamma.

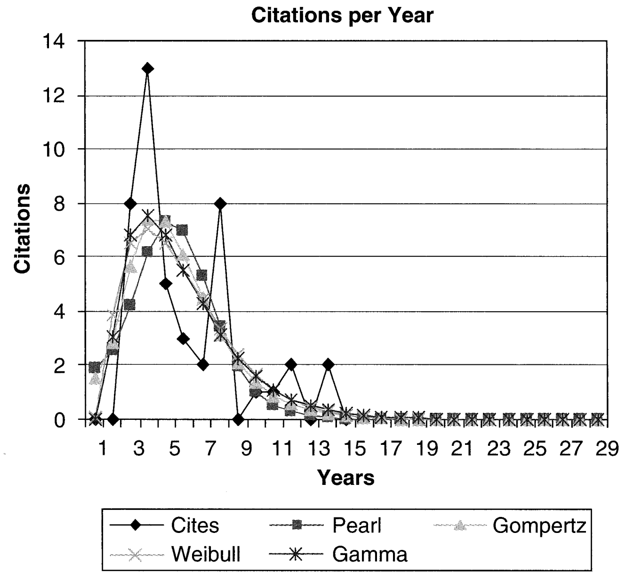

Figure 6 A 'Shooting star' (Jackson (1990)).

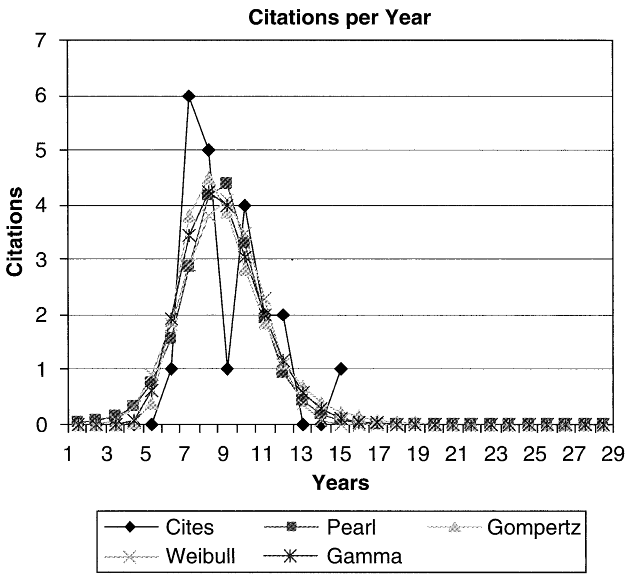

Figure 7 A 'Sleeping beauty' (Dekok (1990)).

even have negative skew, that is, citations peaking later rather than earlier.

The main points emerging from the analysis are:

1. The first group of papers were those that were symmetrical around the point of inflection—the rise and subsequent fall occurred over equal time periods. These were generally fitted best by the Pearl function (six examples). Series with some positive skew could be fitted with any of the other three functions: Gompertz (eight), gamma (five), Weibull (three). Series with a greater degree of skew were best fitted by the Weibull (five) and the gamma (five). Three series had very little pattern with no build up and decline but simply random numbers of citations. These were fitted (poorly) by gamma and Weibull functions with quite extreme parameters.
   Overall (research question 1), these results do not support Burrell's (2002) assumption that all papers within a collection will have the same obsolescence function. Having said that, the gamma and Weibull are very flexible and were not that much worse than the Pearl and Gompertz curves even for symmetrical patterns.
2. In terms of specific patterns (research question 2) some shooting stars and sleeping beauties were identified. Shooting stars will have high initial citations but these will tail off quickly. In one example, a paper had acquired 30 cites in 4 years—7.5 per year—but only gained another 15 in the remaining 10 years—1.5 per year (Figure 6). In terms of the Weibull, such papers would have low values of the scale parameter (<5) with moderate values of shape (1.5–3). Two such papers were Wei(4.83,2.13) and Wei(4.69,1.55) (using the notation Wei(scale,shape)), which were the two lowest scale parameter values in the sample.
3. It was harder to find sleeping beauties given the restrictions on the time period of the sample. Van Raan (2004) characterizes such papers in terms of the sleeping period (typically between 5 and 10 years), the 'depth of sleep' (a 'deep sleep' would average no more than one citation per year), and the 'awake intensity' (average citations after the sleeping period). Such papers would have higher values of the scale parameter (>5) to allow for the sleeping period, together with higher values of the shape parameter (>4). One example is shown in Figure 7, which had only seven cites in its first 6 years but then had 14 in its next 5. This was fitted as Wei(7.87,4.04). Burrell (2005) provides an interesting analysis of the likelihood of sleeping beauties occurring by chance given an underlying gamma-Poisson production process.

Table 5 Total Citations per year for the 36 completed papers

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Citations	23	57	83	79	83	94	101	61	69	54	74	43	19	15

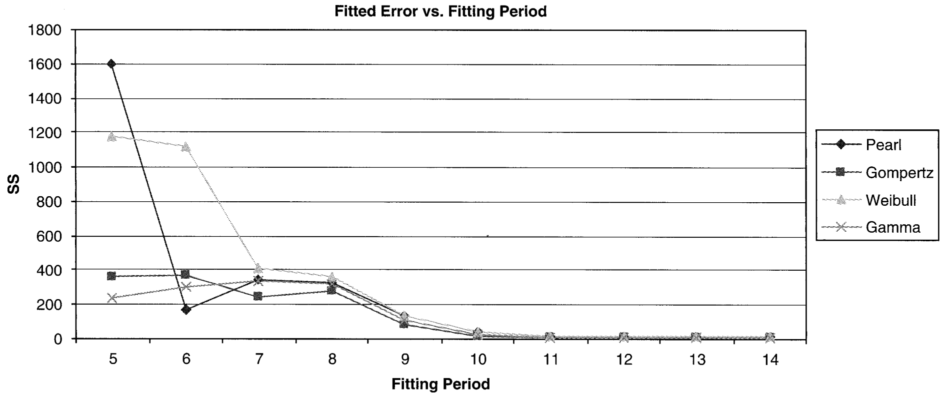

Figure 8 Fitted errors *versus* fitting period for different functions (Dekok (1990)).

4. Finally, one pattern emerged that was not anticipated. Of the 36 examples, nine had significant dips in the number of citations after 8 or 9 years. This is, in fact, illustrated in both of the papers in Figures 5 and 6. The dips often went down to zero or one citation before picking up at close to the previous level. This can be seen clearly in Table 5, which shows the total citations for the sample of 36 papers. This rises to a peak of 101 after 7 years but then falls suddenly in years 8, 9, and 10 before picking up again. The frequency of occurrence and size of this pattern makes it unlikely to have occurred by chance. The most plausible explanation is that initially citations are generated by the first publication of the paper, delayed by the process of writing the citing paper and getting it published. Citations then fall off before a second wave emerges triggered by the later citations rather than the original publication. Although, if this is the explanation, it might have been expected to occur earlier, perhaps after 5 or 6 years. More detailed work would be needed to test this hypothesis.

3.3. *Predicting future numbers of citations*

The next research question to be considered was whether it was possible at all to predict the future number of citations for a single paper based on the pattern of citations in the early years?

To begin with, only the completed papers were considered for here it was possible to know with some certainty what the actual outcome was. This, of course, is a biased sample with respect to the general population as they will all have completed in a relatively short time and will not have the long positive skew of those still active. The method used was based on the above where all 14 years of data were used to fit the curve. Now, we began by using only the first 5 years points to fit the curve as though we were trying to predict the future citations after only 5 years. Thus, SS for all 15 points was recorded. This process was then repeated for 6 years, 7 years and so on up to the full 14. As more years were used for the fitting period, the SS generally fell. This was done for each of the four functions.

The results were, to say the least, highly varied on several dimensions: (i) the SS varied hugely depending on the number of fitting periods; (ii) it also varied significantly between the different functions, especially when there were few fitting periods; but (iii) this did not remain consistent as particular functions rose or fell through the process. Figure 8 shows a typical example for one of the papers. In this case, at the start the gamma and Gompertz were best with SS between 200 and 400, while the Pearl logistic was up at 1600. However, within one period the Pearl changed dramatically to become best with less than 200. In period 7 the Gompertz became best and remained so till the end with a final SS of 7.8. The fit was good from period 10 onwards by which time the SS was under 16.

Taking firstly the overall quality of fit, we can look at the ratio between the period 5 SS and the period 14 SS as a measure of how much worse the fit becomes with less fitting periods. We can get a lower bound on this by taking the best fitting function at each of the two periods. In practice, if we were trying to make predictions *ex ante*, that is, without knowing the future citations, results would be worse than this because we would not know which function would turn out to be best. Thus for Figure 8, the best fit SS using all 14 periods was 7.8, but if only the first five periods were used this rose to 237.9, giving a ratio of 30.5.

Table 6 Summary statistics for SS and years to reasonable fit

	SS: Ratio of 5-per. to 14-per.	*Years to 'reasonable' fit*
Min	1.09	5
Mean (Mode)	29.05	8.9 (10)
Max	256.36	12

Across the papers the mean ratio was 29 times but this rose to as much as 256 times (4633/18) in the worst case. That is, on average the SS based on five periods fitting was 29 times that based on the full 14 periods. These are very large figures especially given that they are based on the best fitting curve at each period (Table 6).

At what period, typically, do the fits become reasonable? The term reasonable is somewhat arbitrary but we took the view that when the SS had come down (from on average 29 times) to only being double the final value it was 'reasonably' close. In Figure 8, for example, this occurred at year 10. This was recorded for each series and from Table 6 we can see that the mean year at which this occurred was 8.9 with the modal value being year 10. In other words, it was only after 10 years worth of data were available for fitting that the SS came to within 100% of the final 14 year value. So the conclusion is that for these data, which are in any case biased in having citation history completed by year 14, the fits are not generally at all reliable before years 9 or 10 by which time there are not many years left anyway. Certainly with reference to the RAE it shows that judging future citations on the basis of citations so far (within the period 2001–2007) is perilous.

It had been hypothesized that the time at which the fits became more reliable may be related to the point of inflection of the cumulative curve (Meade and Islam, 1998). Before this point, it is difficult to decide when the citations are likely to begin to slow down, but after this point, especially a few years later when the downward slope has become established, one might expect that the fitted curves should settle down. The point of inflection is equivalent to the mode of the corresponding pdf. It is unreliable to estimate this empirically from these data since quite commonly a series may have several modes, that is, several periods with the same maximum number of citations, so the inflection points were estimated theoretically from the fitted curves given the formulae in Appendix B. The values estimated from the Weibull and gamma parameters were generally extremely similar ($r = 0.996$) and they were also highly correlated with the Pearl and Gompertz ($r = 0.93$) figures (after removing four unusual observations where the fitted Weibull and gamma shape parameters were < 1—see Appendix B). It turned out that there was, in fact, no correlation between the estimated inflection times and the reasonable fitting period as defined above, nor could regression establish a significant relationship.

Looking next at the best fitting function over time, it had been hypothesized that several patterns might occur. For example, that certain functions would predominate in the early periods (e.g, the Pearl/Gompertz because of their relative symmetry) and different ones later on; that the best fitting function would remain fairly constant over the periods; or that the Weibull/gamma functions would predominate because of their flexibility. In the event none of these were observed. There were no cases where the same function remained best throughout the period and it often changed three or more times. The most common ones in the first period (year 5) were Gompertz (12) and gamma (15) while the most common at the end was the gamma (13). In only a third of cases were the first and final ones the same function even accepting changes in between.

The general conclusions from this section are that it is extremely difficult to predict accurately future citations for an individual paper, at least until it is well through its citation lifetime. And that there are no underlying patterns to the sequence of functions fitted or the levels of errors that are generated.

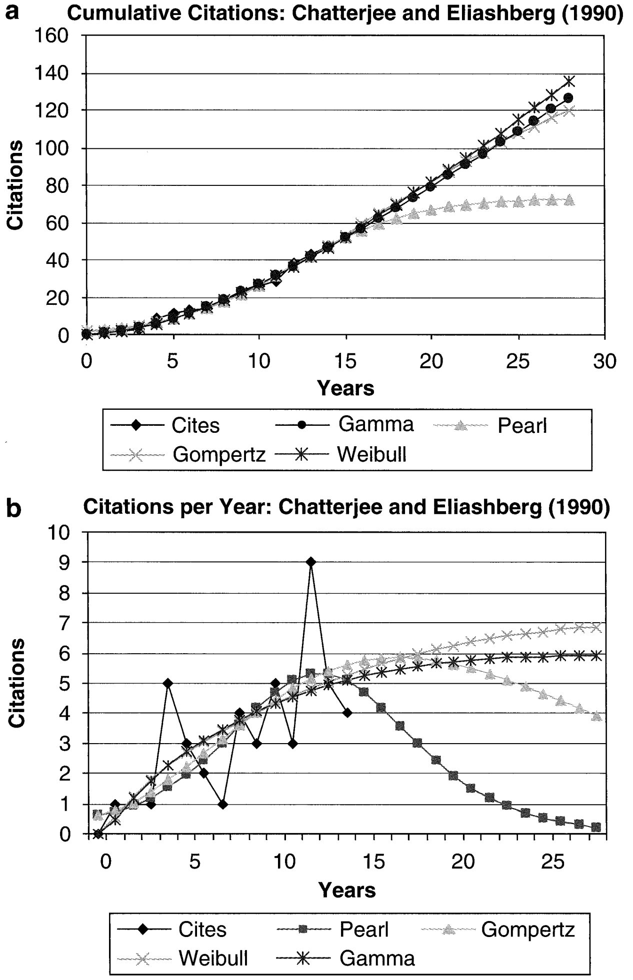

Figure 9 (a) Cumulative citations and fitted curves. (b) Year-on-year citations (equivalent to a pdf).

3.4. Fitting still active papers

Finally, we revisited the problem of fitting papers that were still active. The problems can be illustrated in Figures 9a and b which show a fairly typical paper that is still actively cited.

In Figure 9a the cumulative s-curve seems well-behaved and appears well-fitted by the curves. It is noticeable that the Pearl-logistic gives a much lower limit than the other three. However, the situation is shown more clearly when we look at the year-on-year citations shown in Figure 9b.

We can see that citations rose to a peak in year 4 before falling away to year 7. At this point it would have appeared that the citations were ending but they then pick up significantly reaching another peak in year 12. This dip is common and was discussed above. The next 2 years then fall off again. The question is, what happens next? A further fall would suggest that the point of inflection had passed but it seems equally possible, given the variability of the series, that there could be a rise.

This uncertainty is reflected in the curves that have been fitted. The Pearl, which is symmetrical, turns over and gives the lowest forecast limit of 73. It is always the case with the Pearl on still-active data that it treats the latest peak as the inflection point. The Gompertz gives considerably higher forecasts with a limit of 168. Again its shape is quite constrained. The Weibull and gamma suggest that the citations will carry on rising far into the future and estimate absurd upper limits of 6800 and 4500, respectively. As it turned out, the actual number of citations in the next year (2005) was only one! While this example would appear to strongly favour the Pearl or Gompertz curves, other examples can be found where they significantly underestimate future citations, even when the inflection point has clearly been reached.

The difficulty of predicting future citations demonstrated by this example reflects very well Meade and Islam's (1998, p. 1116) comments about technological forecasting using s-curves:

> [i]t is easy to see how difficult it is to recognise that the point of inflection has been reached. It is even more difficult to predict the future path of the curve. The super-imposition of random noise, the case in practice, serves to make the task of forecasting ... even more demanding.

4. Conclusions

This paper set out to answer three questions concerning the behaviour of citations: (i) to what extent can the citations from collections of papers be modelled by the same obsolescence function? (ii) Can we identify different patterns of citation behaviour and explain them? and (iii) can we predict the number of future citations given the pattern of citations in the first few years?

Looking at citations for collections of papers, the significant lack of symmetry over time meant that they were best fitted by the gamma distribution, and occasionally the Weibull, but not the Pearl logistic or Gompertz curves. This confirms earlier research. The fitted parameters were significantly different between journals reflecting the large disparity in the numbers of citations received.

The move to individual papers brought in a large amount of randomness and variability. Initially analysis concentrated on papers that had generally completed their citation lives to avoid the problems of having to estimate the upper limits to numbers of citations. Here the results were mixed. The gamma and Weibull were best for a majority of papers but the Pearl and Gompertz were best for those which were more symmetrical. This is perhaps surprising as the gamma and Weibull are also capable of taking on symmetrical shapes. This does mean that Burrell's assumption about homogeneity of the obsolescence function is not borne out.

Several patterns were observed in the data including sleeping beauties and shooting stars, and these can be identified through the fitted parameters. An unexpected, but very common, pattern was also observed – that is a dip in citations after 8 or 9 years. This may be due to a shift from citations based on the original paper to those based on other citations but this needs further investigation.

Predicting future citations for individual papers proved to be extremely difficult. For papers whose citations were complete (within the 14 years) the fit became reasonable only after about 10 years, well past the point of inflection. For papers that were still active (the majority) different curves generated wildly different estimates of the potential upper limits.

Finally, it was surprising how many papers were still being actively cited after 14 years. It would be useful to replicate this analysis on a sample that is as old as possible—in the case of ISI data this would be back to 1975.

Acknowledgements—I acknowledge the work of Hajir Karbassi in rigorously collecting the data.

References

Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Automatic Control* **AC-19**: 716–723.

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

Bettancourt L and Houston M (2001). The impact of article method type and subject area on article citations and reference diversity. *Market Lett* **12**: 327–340.

Burrell QL (1990). Using the gamma-Poisson model to predict library circulation. *J Am Soc Inform Sci* **41**: 164–170.

Burrell QL (2001). Stochastic modelling of the first-citation distribution. *Scientometrics* **52**: 3–12.

Burrell QL (2002). The *n*th-citation distribution and obsolescence. *Scientometrics* **53**: 309–323.

Burrell QL (2003). Predicting future citation behaviour. *J Am Soc Inform Sci* **54**: 372–378.

Burrell QL (2005). Are 'sleeping beauties' to be expected? *Scientometrics* **65**: 381–389.

Callahan M, Wears R and Weber E (2002). Journal prestige, publication, bias and other characteristics associated with citation

of published studies in peer-reviewed journals. *J Am Med Assoc* **287**: 2847–2850.
Chatterjee R and Eliashberg J (1990). The innovation-diffusion process in a heterogeneous population – a micromodeling approach. *Mngt Sci* **36**: 1057–1079.
Cunningham S and Bocock D (1995). Obsolescence of computing literature. *Scientometrics* **34**: 255–262.
Dekok A (1990). Hierarchical production planning for consumer goods. *Eur J Opl Res* **45**: 55–69.
DfES (2006). *Reform of higher education research assessment and funding*. DRN05/13, Department for Education and Skills, London.
Doyle J and Arthurs A (1995). Judging the quality of research in business schools. *Omega, Int J Mngt Sci* **23**: 257–270.
Doyle J, Arthurs A, Green R, McAulay L, Pitt M and Bottomly P (1996). The judge, the model of the judge, and the model of the judged as judge: analysis of the UK 1992 Research Assessment Exercise data for Business and Management Studies. *Omega* **24**: 13–28.
Doyle J, Arthurs A, McAulay L and Osbourne P (1996). Citation as effortful voting: a reply to Jones, Brinn and Pendlebury. *Omega* **24**: 603–606.
DuBois FL and Reeb D (2000). Ranking the international business journals. *J Int Bus Studies* **31**: 689–704.
Egghe L (2000). A heuristic study of the first citation distribution. *Scientometrics* **48**: 345–359.
Garfield E (1972). Citation analysis as a tool in journal evaluation. *Science* **178**: 471–479.
Glänzel W and Moed HK (2002). Journal impact measures in bibliometric research. *Scientometrics* **53**: 171–193.
Glänzel W and Schoepflin U (1995). A bibliometric study of ageing and reception processes of scientific literature. *J Inform Sci* **21**: 37–53.
Harzing, A.-W. (2005). *Journal Quality List*, http://www.harzing.com/.
Jackson M (1990). Beyond a system of systems methodologies. *J Opl Res Soc* **41**: 657–668.
Jennings C (1998). Citation data: the wrong impact? *Nat Neurosci* **1**: 641–642.
Jones M, Brinn T and Pendlebury M (1996). Journal evaluation methodologies: a balanced response. *Omega* **24**: 607–612.
Jones M, Brinn T and Pendlebury M (1996). Judging the quality of research in business schools: a comment from accounting. *Omega Int J Mngt Sci* **24**: 597–602.
MacRoberts M and MacRoberts B (1987). Problems of citation analysis: a critical review. *J Am Soc Inform Sci* **40**: 342–349.
Martino J (1983). *Technological Forecasting for Decision Making*. North-Holland: New York.
Meade N and Islam T (1998). Technological forecasting—model selection, model stability and combining models. *Mngt Sci* **44**: 1115–1130.
Mingers J and Burrell Q (2006). Modelling citation behavior in management science journals. *Inform Process Mngt* **42**: 1451–1464.
Mingers, J. and Harzing, A.-W. (2005). *Ranking journals in business and management: a statistical analysis of the Harzing dataset*. Working Paper 85, Kent Business School, Canterbury.
Pearl R and Reed L (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc Nat Acad Sci* **6**: 275–288.
Redner S (2005). Citation statistics from 110 years of Physical Review. *Phys Today* **58**: 49–56.
Rousseau R (1994). Double exponential models for first citation processes. *Scientometrics* **30**: 213–227.
Schwarz G (1978). Estimating the dimension of a model. *Ann Statist* **6**: 461–464.
Seglen P (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ* **314**: 498–502.
Sharif N and Islam M (1980). The Weibull distribution as a general model for forecasting technological change. *Technol Forecasting Soc Change* **18**: 247–256.
Stone R (1978). Sigmoids. *Bull Appl Statist* **7**: 59–119.
Tahai A and Meyer M (1999). A revealed preference study of management journals' direct influences. *Strategic Mngt J* **20**: 279–296.
van Dalen H and Henkens K (2001). What makes a scientific article influential? The case of demographers. *Scientometrics* **50**: 455–482.
Van Raan AJ (2004). Sleeping beauties in science. *Scientometrics* **59**: 467–472.

Appendix A. The Gamma-Poisson model with obsolescence

We assume a collection of papers, all from the same year, with some common characteristic such as publishing journal. Each paper is assumed to generate citations at a constant latent rate (λ) following the Poisson distribution but these rates vary across the collection as a random variable Λ.

Then, the probability that a paper with latent rate λ will generate r citations by time t is (Burrell, 2001, 2002)

$$P(X_t = r|\Lambda = \lambda) = \frac{e^{-\lambda t}(\lambda t)^r}{r!} \tag{A.1}$$

The population distribution (ie, for a randomly chosen paper of unknown latent rate) will be a *mixture* of the above distributions according to the distribution of rates:

$$P(X_t = r|\Lambda) = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^r}{r!}\, dF_\Lambda(\lambda) \tag{A.2}$$

where $F_\Lambda(\lambda)$ is the cumulative distribution function of the latent rate known as the mixing distribution.

If we assume that the mixing distribution is gamma as follows:

$$dF_\Lambda(\lambda) = f_\Lambda(\lambda) = \frac{\alpha^\nu \lambda^{\nu-1} e^{-\alpha\lambda}}{\Gamma(\nu)}$$
$$\text{(characterized as } \Gamma(\alpha, \nu)) \tag{A.3}$$

(where (ν) is a shape parameter and (α) is a size/scale one) then it can be shown that the resulting distribution is the negative binomial:

$$P(X_t = r) = \binom{r+\nu-1}{\nu-1}\left(\frac{\alpha}{\alpha+t}\right)^\nu \times \left(1-\frac{\alpha}{\alpha+t}\right)^r, \quad r=0,1,2,\ldots \tag{A.4}$$

where $E[X_t]=\mu(t)=\nu t/\alpha$; and $\mathrm{Var}(X_t)=\sigma^2(t)=\nu t(t+\alpha)/\alpha^2$.

We can summarize this by writing:

$$X_t \approx \mathrm{NB}\left(\frac{\alpha}{\alpha+t}, \nu\right) \tag{A.4*}$$

We now allow the latent rate to vary over time as well as across papers. Burrell (2001) makes the assumption that while the latent rate will differ between papers, the obsolescence function will be the same for all within a particular collection. This will be of the form $\lambda t = \lambda c(t)$ where $c(t)$ reflects the changing pattern of λ over time.

This gives the following generalization of (1) for a paper with known λ:

$$P(X_t = r|\Lambda = \lambda) = \frac{e^{-\lambda C(t)}(\lambda C(t))^r}{r!} \tag{A.1a}$$

where $C(t)$ is the cumulative of $c(t)$.

And of (2) for a randomly chosen paper:

$$P(X_t = r|\Lambda) = \int_0^{\infty} \frac{e^{-\lambda C(t)}(\lambda C(t))^r}{r!}\, dF_\Lambda(\lambda) \tag{A.2a}$$

If we assume, as we have above, that the latent rate distribution is a gamma, then the gamma-Poisson model becomes (Burrell, 2002):

$$P(X_t = r) = \binom{r+v-1}{v-1}\left(\frac{\alpha}{\alpha + C(t)}\right)^v \times\left(1-\frac{\alpha}{\alpha+C(t)}\right)^r, \quad r=0, 1, 2, \ldots \tag{A.4a}$$

Thus we have that

$$X_t \approx \mathrm{NB}\left(\frac{\alpha}{\alpha + C(t)}, v\right) \tag{A.4*a}$$

It is the above assumption concerning the obsolescence function $C(t)$ and its possible forms that are explored in this paper.

Appendix B. S-curves used in the study

B.1. The logistic curve (Pearl-Reed curve)

Probably the most widely known growth curve was developed originally by Verhulst in 1838 and then popularized by geographers Pearl and Reed (1920). The underlying assumption is that initially the rate of growth is proportional to the size of the population, but that as size increases environmental restrictions will reduce growth until saturation is reached. The derivation and various formulations are explained in Stone (1978).

Equation

$$Y_t = \frac{L}{(1 + ae^{-bt})} \tag{B.1}$$

Parameters

- L upper limit
- a scale parameter affecting the location of the curve
- b shape parameter affecting the steepness/shape of the curve

Note that a and b are independent in that changes in location do not affect the shape.

Characteristics

The curve is symmetrical about its point of inflection, which corresponds to the maximum growth rate. This occurs at $Y = L/2$ when $t = \ln(a)/b$

The growth rate is given by

$$Y' = bY\left(\frac{L-Y}{L}\right) \tag{B.1a}$$

which shows that the growth at any point depends both on distance to go $(L - Y)$ and distance travelled (Y).

And the proportionate growth by

$$\frac{Y'}{Y} = b\left(\frac{L-Y}{L}\right) \tag{B.1b}$$

which shows that the proportionate growth is a linear function of the growth so far.

B.2. Gompertz curve

This curve was first formulated by Gompertz in 1825 and differs from the logistic in not being symmetrical about the point of inflection.

Equation

$$Y_t = Le^{-ae^{-bt}} \tag{B.2}$$

Parameters

- L upper limit
- a scale parameter affecting the location of the curve
- b shape parameter affecting the steepness/shape of the curve

Characteristics

The inflection point is where $Y = L/e$ when $t = \ln(a)/b$. It thus occurs at the same time as the logistic but the growth value is less—only 73%. Growth is steeper before the inflection point than after it.

The growth rate is given by

$$Y' = bY \ln\left(\frac{L}{Y}\right) \tag{B.2a}$$

which, for large Y, can be approximated by $Y' = b(L - Y)$ showing that for later period growth depends only on distance to go to the upper limit, not on previous history.

The proportionate growth is

$$\frac{Y'}{Y} = b\ln\left(\frac{L}{Y}\right) \tag{B.2b}$$

showing that proportionate growth is not linear but reduces as Y approaches the limit.

B.3. Weibull distribution

The Weibull is a statistical distribution commonly used in reliability studies. It was suggested that the cumulative probability distribution (CDF) could be used as an s-curve by Sharif and Islam (1980). It is a very flexible distribution whose probability function can take a variety of shapes from right skew through normality to left skew. It is non-symmetric and flexible in Meade and Islam's (1998) terms.

Equation

The Weibull cdf is given by

$$Y_t = L(1 - \mathrm{e}^{-(t/a)^b}) \tag{B.3}$$

Parameters

- L upper limit
- a scale parameter affecting the location of the curve
- b shape parameter affecting the steepness/shape of the curve

Characteristics

The inflection point is when $t = a(1 - 1/b)^{1/b}$, where $Y = L(1 - \mathrm{e}^{-(1-1/b)})$. (Note that for $b < 1$ the formula breaks down and t is defined as 0 which is the modal point of the probability distribution). This is in contrast to the previous curves which had constant values of Y. For the Weibull the value of Y at inflection depends on the parameter value b. This provides a greater degree of flexibility in modelling the point of decline in citations.

The growth rate is given by

$$Y' = \frac{b}{a}(L - Y)\ln\left(\frac{L}{L-Y}\right)^{1-1/b} \tag{B.3a}$$

which shows that growth depends only on the distance to the upper limit.

The proportionate growth is

$$\frac{Y'}{Y} = \frac{b}{a}\left(\frac{L-Y}{Y}\right)\ln\left(\frac{L}{L-Y}\right)^{1-1/b}$$

Showing that proportionate growth is non-linear.

B.4. Gamma distribution

Another very flexible probability distribution used extensively in queuing and waiting situations. It is similar to thet Weibull in taking a variety of shapes from the exponential to the normal. It is non-symmetric and flexible in Meade and Islam's (1998) terms.

Equation

The gamma cdf is given by

$$Y_t = L\frac{\gamma(b, t/a)}{\Gamma(b)} \tag{B.4}$$

where $\Gamma()$ is the complete gamma function and $\gamma()$ is the incomplete gamma function. This is an awkward equation form and it is more usually seen as a pdf:

$$y_t = \frac{a^{-b}t^{b-1}\mathrm{e}^{-t/a}}{\Gamma(b)}$$

Parameters

- L upper limit
- a scale parameter affecting the location of the curve
- b shape parameter affecting the steepness/shape of the curve

Characteristics

The inflection point is when $t = a(b - 1)$. (For $b < 1$ the value is defined to be 0.) This corresponds to a Y value of $Y_t = L[\gamma(b, b-1)/\Gamma(b)]$ but because of the nature of the cumulative gamma function there is no easy expression for this value. It can be calculated numerically. As with the Weibull, the value of Y at inflection depends on the parameter value b. It is also difficult to formulate expressions for the growth rate.

Received November 2005;
accepted March 2007 after one revision