Made available by Hasselt University Library in https://documentserver.uhasselt.be

A model showing the increase in time of the average and median reference age and the decrease in time of the Price index

Peer-reviewed author version

EGGHE, Leo (2010) A model showing the increase in time of the average and median reference age and the decrease in time of the Price index. In: SCIENTOMETRICS, 82(2). p. 243-248.

DOI: 10.1007/s11192-009-0057-3
Handle: http://hdl.handle.net/1942/9283

A model showing the increase in time of the average and median reference age and the decrease in time of the Price Index

by

L. Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium[1]
and
Universiteit Antwerpen (UA), Campus Drie Eiken, Universiteitsplein 1, B-2610 Wilrijk, Belgium

leo.egghe@uhasselt.be

ABSTRACT

This paper proves two regularities that where found in the paper [V. Larivière, E. Archambault and Y. Gingras (2007). Long-term patterns in the aging of the scientific literature, 1900-2004. Proceedings of ISSI 2007. CSIC, Madrid, Spain, 449-456, 2007]. The first is that the mean as well as the median reference age increases in time. The second is that the Price Index decreases in time.

---

[1] Permanent address
Key words and phrases: mean reference time, median reference time, Price Index, exponential growth.
Acknowledgement. The author is grateful for interesting discussions with Prof. Dr. R. Rousseau on the topic of this paper.

Using an exponential literature growth model we prove both regularities. Hence we show that the two results do not have a special informetric reason but that they are just a mathematical consequence of a widely accepted simple literature growth model.

I. Introduction

The interesting paper Larivière, Archambault and Gingras (2007) – see also Larivière, Archambault and Gingras (2008) - reveals that, using the extensive Thomson Scientific databases (all fields), the average (i.e. mean) reference age is an increasing function of time (sixties until now). Stated more clearly, as the present time passes, if we calculate the average reference age, we get an increasing function of time. This is even so when one limits the cited period to 20 or even 100 years: the more the present time goes forward the longer the effective cited period, hence, in this paper we will not restrict the cited period. The same can be said about the median reference age. The authors call this a "counter-intuitive phenomenon" and, at first sight, they are right: one always thinks that, as the sciences grow, authors use more and more recent articles in their reference lists. But one may not forget that growing sets of recent literature become older when time passes so that there are also more and more older articles that can be cited. The outcome of this double effect on the age of references is not clear intuitively and needs further investigation. This is the topic of this paper.

The same authors also found that the Price Index (PI) is decreasing in time since the mid fifties. The PI is the fraction of references (e.g. for a given year) that are 5 years old or younger. In fact, one does not have to take 5 as the time span: any number $d = 1, 2, 3, ...$ can be used. So does one use $d = 2$ in Glänzel and Schoepflin (1995).

The two results above lead to the (this time intuitively clear) conclusion that PI is decreasing with the mean reference age and the same is true for the median reference age. This (weaker) conclusion (it does not imply that the average and median reference age increases in time and that PI decreases in time) was already found experimentally in Glänzel and Schoepflin (1995) and explained in Egghe (1997) using an exponential aging model. Using such a model we found the average reference age (denoted MA) and the median reference age (denoted MD) as a function of the parameters of the aging model. The same was done for PI, which showed then the decreasing relation MA ® PI and MD ® PI. In this argument, hence, no time-dependent model for $MA(t)$, $MD(t)$ and $PI(t)$ was developed.

In order to be able to explain the two results found in Larivière, Archambault and Gingras (2007), we hence have to make a time-dependent model for $MA(t)$, $MD(t)$ and $PI(t)$. This is done in this paper.

II. The model

II.1. Introduction

Our model for the time-dependent description of $MA(t)$, $MD(t)$ and $PI(t)$ is based on the simple exponential model for the growth of literature. There are, basically, two different growth types: the exponential type and the S-shaped type – see also Egghe and Rao (1992). These apply to various phenomena in life: physical growth of living beings, of economic features, of literature or of the Internet.

Most classically – and simplifying – one can say that young phenomena start growing exponentially and that, at a certain point in time, the S-shape starts. In this sense we can say that, at this point in time, the Internet is still "young": the encountered yearly growth curves are still exponential – see e.g. Adamic and Huberman (2001), Internet Software Consortium (2007). Even when growth starts to be S-shaped, we can consider the exponential model as a first approximation.

Somewhat surprisingly, the Science Citation Index (SCI) also does not show an S-shape for the yearly growth. We refer here to Jin and Rousseau (2005) showing that an exponential model of the type ($g(t)=$ number of articles at time t)

$$g(t)= g_0 g^t \tag{1}$$

for $g > 1$ (but close to 1) can be fitted. The exponential model (1) also approximates growth in case we have an S-shape. For all these reasons we adopt (1) as our growth model.

Next we assume that the probability for a reference to be of year t is proportional to the number (or fraction) $g(t)$ of publications in year t. This assumption is certainly logical since we are dealing with large document sets, hence each having the same distribution of lowly and highly cited documents.

It is the more interesting to study Larivière et al's finding in the framework assuming that there is an exponential decay of references that are s years old $(g(t-s), s \in [0,t])$, an assumption which is classical and called "exponential ageing (obsolescence)" and which <u>seems</u> to contradict Larivière et al's finding (but this is not the case !).

Based on these simple assumptions, we can model the average reference age $MA(t)$, the median reference age $MD(t)$ and the Price Index $PI(t)$ in function of time.

II.2 The function MA(t)

As explained in subsection II.1, we will treat (1) as the number of references that were published in year t (but t will be treated as a continuous variable). Then, by definition of $MA(t)$ we have:

$$MA(t) = \frac{\int_0^t s g(t-s) ds}{\int_0^t g(t-s) ds} \tag{2}$$

assuming that the same proportionality holds if we fix the citing year t (note that here g_0 is a function of t but, as will be clear from the sequel, g_0 cancels in the calculations). Applying (1) now gives

$$MA(t) = \frac{\int_0^t s g_0 g^{t-s} ds}{\int_0^t g_0 g^{t-s} ds}$$

$$MA(t)=\frac{\int_0^t \frac{s}{\gamma^s}ds}{\int_0^t \frac{ds}{\gamma^s}} \tag{3}$$

In order to prove that $MA(t)$ increases in t, we calculate its derivative

$$MA'(t)=\frac{\left(\int_0^t \frac{ds}{\gamma^s}\right)\frac{t}{\gamma^t}-\left(\int_0^t \frac{s}{\gamma^s}ds\right)\frac{1}{\gamma^t}}{\left(\int_0^t \frac{ds}{\gamma^s}\right)^2}$$

Hence $MA'(t)>0$ if and only if

$$t\int_0^t \frac{ds}{\gamma^s}>\int_0^t s\frac{ds}{\gamma^s}$$

which is satisfied since $s\in[0,t]$.

So we have proved that MA is an increasing function of time t. We will now prove the same result for MD.

II.3 The function MD(t)

Using the same functions as above we have that the defining equation for $MD(t)$ is

$$\int_0^{t-MD(t)} g(t-s)ds=\frac{1}{2}\int_0^t g(t-s)ds$$

or, simpler:

$$\int_0^{MD(t)} g(s)ds=\frac{1}{2}\int_0^t g(s)ds$$

Applying (1) once again (note that g_0 (dependent on t) cancels) we find

$$\int_0^{MD(t)} g_0 g^s ds = \frac{1}{2}\int_0^t g_0 g^s ds$$

or

$$MD(t) = \frac{1}{\ln g}\ln\left(\frac{1}{2}g^t + \frac{1}{2}\right) \tag{4}$$

which is clearly increasing in t since $g > 1$.

Finally, we will show that $PI(t)$ decreases in t.

II.4 The function PI(t)

Let us study the Price Index being the fraction of references that are less than or equal to d years old ($d > 0$, continuous variable). Since this Price Index, denoted $PI = PI(t)$, is dependent on t we will omit the dependence notation on d since d is fixed here. Hence

$$PI(t) = \frac{\int_0^d g(t-s)ds}{\int_0^t g(t-s)ds} \tag{5}$$

Using (1) again we easily find

$$PI(t) = \frac{g^t - g^{t-d}}{g^t - 1} \tag{6}$$

$$= \left(1 - g^{-d}\right)\frac{g^t}{g^t - 1}$$

$$= \left(1 - g^{-d}\right)\left(1 + \frac{1}{g^t - 1}\right) \tag{7}$$

which is clearly decreasing in t since $g > 1$.

It is, furthermore, easy to show that $PI(t)$ is a convexly decreasing function of t. Its limit, for $t \to \infty$ is interesting. Clearly from (7) we have

$$\lim_{t \to \infty} PI(t) = 1 - \frac{1}{g^d} \tag{8}$$

Formulae (6) and (8) also show that $PI(t)$ increases in d, a logical fact. Formula (8) is a continuous version of the discrete formula, proved in Egghe (1997), where $t = \infty$ was used (time-independent model).

Formula (8) is interesting since it is, for large t, the simple formula for the Price Index with time window of length d. So it applies to citation databases dealing with large time-periods such as Web of Science (WoS) or Scopus.

As calculated in Jin and Rousseau (2007), g for the Science Citation Index (SCI) can be estimated to be

$$g = e^{0.0326} = 1.0331372$$

or about 3.3% increase per year. Using (8) we find for the Price Index PI, for $t \approx \infty$ (i.e. t large) the values as in table 1.

Table 1 PI in function of d for SCI

d	PI
1	0.0320743
2	0.0631199
3	0.0931697
4	0.1222557
5	0.1504088
6	0.1776589
7	0.2040349

It is clear from Table 1 that the increase is quasi linear in d.

This heuristic finding is also seen from (8) as follows: denoting $g = e^{\lambda}$ we have by (8):

$$1 - \frac{1}{g^d} = 1 - e^{-\lambda d}$$

$$= 1 - \left(1 - \lambda d + \frac{(\lambda d)^2}{2} - \ldots\right)$$

$$= \lambda d\left(1 - \frac{\lambda d}{2} + \ldots\right)$$

which shows the quasi – linearity in Table 1 since λ is small (This argument was proposed to me by one of the referees to whom my sincerest thanks).

Note: Also in the poster paper Chen, Liu and Liang (2007) one has found experimentally that $\mathrm{Md}(t)$ increases in t and that $\mathrm{PI}(t)$ (for different values of d) decreases in t. Here 106 journals about mechanics in WoS are used.

III. Comments and conclusions

Using an exponential growth model for the number of references in a certain year we proved, as a mathematical consequence of this model, that the average and median reference age is an increasing function of time. Hence, this at first surprising result, confirmed by data in Larivière, Archambault and Gingras (2007) and in Chen, Liu and Liming (2007), has no specific (direct) informetric explanation but is a mathematical consequence of a simple, well-known, exponential growth model.

In the same way we prove that all Price Indices are decreasing in time, again confirmed in the above mentioned papers. Again this is a mathematical consequence of the exponential growth model and has no specific (direct) informetric background, i.e. there is no specific citation behavior that explains these findings.

The consequence of these two results is that the Price Index is a decreasing function of the average and median reference age, which is not a surprising result. This result was already proved in Egghe (1997) in a time-independent context and based on experimental data in Glänzel and Schoepflin (1995).

References

Adamic L.A. and Huberman B.A. (2001). The Web's hidden order. Communications of the ACM 44(9), 55-60.

Chen L., Liu Z. and Liang L. (2007). The relationship of Price Index and median citation age. In: Proceedings of ISSI 2007 (CSIC, Madrid, Spain, D. Torres-Salinas and H.F. Moed (eds.)), 844-845.

Egghe L. (1997). Price index and its relation to the mean and median reference age. Journal of the American Society for Information Science 48(6), 564-573.

Glänzel W. and Schoepflin U. (1995). A bibliometric ageing study based on serial and non-serial reference literature in the sciences. Proceedings of the Fifth Biennial Conference of the ISSI (Rosary College, River Forest, Ill., USA, M.E.D. Koenig and A. Bookstein (eds.)), 177-185.

Internet Software Consortium (2007). Internet Domain Survey. Host Count. http://www.isc.org/index.pl?/ops/ds/hosts.php

Jin B. and Rousseau R. (2005). China's quantitative expansion phase: exponential growth but low impact. In: Proceedings of ISSI 2005 (Karolinska University Press, Stockholm, Sweden, P. Ingwersen and B. Larsen (eds.)), 362-370.

Larivière V., Archambault E. and Gingras Y. (2007). Long-term patterns in the aging of the scientific literature, 1900-2004. In: Proceedings of ISSI 2007 (CSIC, Madrid, Spain, D. Torres-Salinas and H.F. Moed (eds.)), 449-456.

Larivière V., Archambault E. and Gingras Y. (2008). Long-term variations in the aging of scientific literature: from exponential growth to steady state science (1900-2004). Journal of the American Society for Information Science and technology 59(2), 288-296.