Check for updates

# A model for reference list length of scholarly articles

Fatemeh Ghaffari[1,2] · Mark C. Wilson[1,2]

## Abstract

We introduce and analyse a simple probabilistic model of article production and citation behavior that explicitly assumes that there is no decline in citability of a given article over time. It makes predictions about the number and age of items appearing in the reference list of an article. The latter topics have been studied before, but only in the context of data, and to our knowledge no models have been presented. We then perform large-scale analyses of reference list length for a variety of academic disciplines. The results show that our simple model cannot be rejected, and indeed fits the aggregated data on reference lists rather well. Over the last few decades, the relationship between total publications and mean reference list length is linear to a high level of accuracy. Although our model is clearly an over-simplification, it will likely prove useful for further modeling of the scholarly literature. Finally, we connect our work to the large literature on "aging" or "obsolescence" of scholarly publications, and argue that the importance of that area of research is no longer clear, while much of the existing literature is confused and confusing.

**Keywords** Bibliometrics · Scientometrics · Growth of scientific literature

## Introduction

The growth of science (which we take as a synonym for "published scholarly findings", including in fields such as humanities and social sciences) can be measured by the number of articles published, the number of authors, the number of cited articles, etc. These have been heavily studied, and each is a "global" indicator that requires complete enumeration. A more "local" indicator is the length of reference lists (that is, the number of bibliography items) in an article. The distribution, or simply a statistic such as the mean or median, can be well estimated via sampling a relatively small number of articles. Another convenient and obvious property is that the reference list of a published article remains unchanged over time.

✉ Mark C. Wilson
markwilson@umass.edu

1    Manning College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, USA

2    Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, USA

The overall distribution of the length of reference lists is obviously related to the overall distribution of citations. If we consider the directed graph where each article is a node and there is an edge from *X* to *Y* precisely when *X* includes *Y* in its reference list, then the total outdogree euals the total indegree: in other words, the number of references over all articles equals the total number of citations to all articles. However, the number of citations to a given article increases over time, which makes reference lists more convenient for many computations. For example, details about reference list length have been used for normalization in citation analysis among other topics (Marx & Bornmann, 2016). This distinction between *synchronous* (looking at reference lists) and *diachronous* (looking at citations) bibliometric analysis has been made by many authors since Line & Sandison in the early 1970s. The term *diasynchronous* has also been used to describe time series of synchronous studies.

The length of reference lists has been much less studied than the number of citations received, or the number of articles published. We recommend (Yitzhaki & Ben-Tamar, 1991) for a review of older literature, which was mostly a collection of small-sample statistical data about reference lists in various research fields or particular journals. It is easily noticeable from the older data that papers from many decades ago tended to include fewer references than those written today. More refined analysis shows that the number of references has been increasing, that there is considerable variation between research fields in the mean or median reference list length, and that there is much variation between article types (for example, review articles tend to include more references than original research articles). More recent and detailed data has been compiled (Larivière et al., 2007; Ucar et al., 2014).

## Our contribution

Although several researchers have modeled the growth in the number of published articles, we are not aware of any modeling of the length of reference lists—to our knowledge all previous work has been purely descriptive based on data. We present a simple continuous-time model for article production that links quantities, and explore its behavior. The model explicitly denies any "obsolescence" process by which older articles become less likely to be cited than newer ones. We use a large dataset of articles published in the period 2006–2016 and show that that the model fits remarkably well for our purposes, despite clearly being an oversimplification. In particular, we find that over the period studied, the growth in mean reference list length is well modeled by a linear function, as is the growth of the number of articles produced, while there is a clear linear relationship between the total number of articles and mean reference list length. We also obtain a better fit than expected (although not a good fit) to the distribution of the lengths of reference lists, which is a much more stringent test of the model. This overall good fit to data gives support to the use of the model in other contexts. We also include a generalization of the model which may be useful for more refined modeling.

We connect the above work to the literature on aging and obsolescence, and give a review of some confusions in that literature.

## A basic model

A good general reference for basic material on citations, references, obsolescence, etc, is Cassidy (2018) (although the reader should note that it mixes up the definitions of the terms "synchronous" and "diachronous"). Some of the key factors that seem likely to

increase the length of reference lists, given by previous authors (for example (Yitzhaki & Ben-Tamar, 1991; Ucar et al., 2014)), are:

- As time progresses there are simply more possible works available that a researcher could cite;
- Modern authors use electronic keyword search, making it easier to find relevant references than in the print journal era;
- Page limits have been relaxed because of online publication, allowing more space in which to cite other work;
- Changes in citation practices (for example: peer review is more stringent now, and reviewers often suggest papers to cite; courtesy citations may be more prevalent now).

The main likely countervailing force is the "obsolescence" of scientific articles, which has been studied by many authors. The rapid growth in literature is often accompanied by an improvement in understanding, meaning that older references are completely subsumed by newer ones and no longer cited (certainly, Newton's original work receives very few citations today) (Cassidy, 2018); also, authors may prize novelty and have only finite stamina for literature review. Thus older papers may become less "citable", by which we simply mean less attractive to a citing author than newer ones.

However, obsolescence is not obviously a major factor, owing to the many ways in which researchers find and study potentially citable works. Ellis (1993), in an influential study that develops a model for information-seeking behavior of researchers, discusses among other topics "monitoring" (keeping up to date with a given topic by regular reading of latest findings), and "chaining", in which a researcher reads a given article, then follows up its references, etc. The former would contribute to obsolescence, but the latter would presumably give an advantage to older works, especially if a researcher chooses the most cited items first, since these are more likely to be older references.

It is therefore very unclear what we ought to expect from the time evolution of citability. However, the above considerations lead to two clear research hypotheses. Hypothesis A states that citability of articles is constant over time. In other words, there is no clear bias either toward or away from older articles. Hypothesis B states that citability of articles declines over time, so that there is a bias toward newer articles. Hypothesis B has been assumed (after an initial increase in citability soon after publication) by many authors in the literature on aging/obsolescence of scholarly articles.

## The model

Consider the following model of article production. Suppose that articles are published continuously and let $P(t)$ be the total number of articles published up to time $t$, so that $P'(t)$ is the instantaneous rate of publication at time $t$. Suppose that each new article cites each previously published article with probability $q(a)$ depending only on the age $a$. In other words, the number of citations by article $X$ published at time $t$ of each article $Y$ published at time $s < t$ is a Bernoulli random variable $c_{t-s,XY}$, and for a fixed value $t - s$ these variables are independent and identically distributed.

The simplest special case, which explicitly uses Hypothesis A, is when $q$ is independent of $a$. In other words, the number of citations by article $X$ published at time $t$ of each article $Y$ published at time $< t$ is a Bernoulli random variable $c_{XY}$, these variables are independent and identically distributed, and independent of $t$. We call this "the uniform model" below.

The quantity $q$ in the model should be expected to vary between research fields, even in the simplest case when it is constant. Much research shows disciplinary differences in citation statistics (see, for example, the study (Boyack, 2018) of over 5 million articles). Our data below shows that the number of references varies substantially between disciplines. In our professional (anecdotal) experience, Mathematics articles are typically written fairly tersely, and citations are mostly given to work which is directly built on by the article in question. By contrast, articles in Economics tend to have noticeably longer introductions with many citations. More detailed study of disciplinary differences in author citation behavior, beyond the difference in field size and speed of publication.

## Behavior of the model

Our first prediction is that the expected length $L(t)$ of the reference list of each article published at time $t$ is

$$L(t) = \int_{-\infty}^{t} q(t - s)P'(s)\,ds$$

The intuition behind this is that $P'(s)\,ds$ is the approximate number of papers published in small time interval $[s, s + ds]$, the age for such papers is approximately $t - s$, and so each is cited with probability $q(t - s)$.

Under the uniform model we have the special case

$$L(t) = qP(t)$$

Differentiating, we see that in this case the growth rate of $L$ is proportional to the growth rate of $P$.

It will often be convenient to start time at some $t_0 > -\infty$ (for example, because we only have data since time $t_0$). In this case we can simply rewrite to obtain

$$L(t) = L(t_0) + \int_{t_0}^{t} q(t - s)P'(s)\,ds$$

An important feature of the model is that because we assume uniformly random choice of article to cite, if we consider the restriction whereby we count only citations to articles published at time $t_0$ or later, the analysis is essentially unchanged. For example, defining $L^*(t)$ to mean the number of references published at time $t_0$ or later, and $P^*(t) = P(t) - P(t_0)$ the number of articles published at time $t_0$ or later, we have

$$L^*(t) = \int_{t_0}^{t} q(t - s)(P^*)'(s)\,ds$$

## Experimental setup

We used a dataset supplied by the company Academic Analytics https://aarcresearch.com/. The dataset is rather comprehensive and covers all publications in the period 2006–2016 by faculty members at a list of 390 PhD-granting universities in North America. More details on the process used can be found in Appendix A. We note that such a comprehensive

dataset is not really necessary for much of the analysis in the present article, because random sampling will be good enough.

We filtered out entries with missing data and those with fewer than 5 references (in order to eliminate letters to the editor). The dataset, taken as a sample of the total amount of articles published over the given time period, seems likely (to us) to be unbiased as far as reference list length goes. In this dataset we have (147,149, 53,776, 30,160, 16,957) articles published in 2006–2016 from the fields of Chemistry, Mathematics, Economics and Oncology respectively, and the number of pages and number of references of each article.

We also needed estimates of $P(t)$. For this we used manual queries in Scopus. Of course, we do not know whether the research field classification used by Scopus is the same as that used in our provided dataset. This was one reason why we restricted to the four fields listed above—the original dataset contained several other research fields such as History and Geography, where we expected some classification inconsistencies.

All statistical analyses were performed using Python with standard packages `numpy`, `pandas`, and `matplotlib`.

## Results

In each research field under investigation, we found a steady increase over the period 2006–2016 in the mean number of articles published each year, and in the number of references per article. The fit of each via ordinary linear regression was good (see Table 1), suggesting a linear growth in reference list length over time. We found noticeable and consistent differences by research field in the mean reference list length, as shown in Fig. 1 (right).

We compared the mean list length $L(t)$ against our estimates of $P(t)$ from Scopus. The uniform case of our model predicts an affine relationship (there is a nonzero constant term coming from all articles published up to 2006) of the form $L(t) = q\big(P(t) - P(t_0)\big) + d$. The fit was extremely good (see Fig. 2 and Table 1). We expect that different research cultures will lead to different values of $q$ in our model; for example we would expect the value of $q$ in the model to be larger for Economics than Mathematics, to the extent that they are disjoint fields. The slope values we obtained from the least squares fitting mentioned above indeed differed substantially, and that for Economics was much larger than that for Mathematics.

We performed several robustness checks on the above results. The appendix shows how replacing the mean with the median makes relatively small changes. We also checked against another dataset. Data kindly supplied by the authors of the detailed study Larivière et al. (2007) show clearly that from around 1980 there has been a roughly linear growth in $P$ and $L$, and that $P$ is closely fitted by a linear function of $L$ (see Fig. 3 and Table 2).

**Table 1** Fit of linear models for $L(t)$ and $P(t)$, 2006–2016 (as shown in Figs. 1 and 2)

| Field | $R^2$ for $L(t)$ | $R^2$ for $P(t)$ | $R^2$ for $L$ vs $P$ | $R^2$ for $P(t) - P(t-1)$ |
|---|---|---|---|---|
| Chemistry | 0.405948 | 0.997372 | 0.963810 | 0.964448 |
| Mathematics | 0.605758 | 0.997737 | 0.993011 | 0.947762 |
| Economics | 0.521143 | 0.968208 | 0.978862 | 0.908734 |
| Oncology | 0.269766 | 0.993056 | 0.888816 | 0.969280 |

(A) $L(t)$ vs $t$



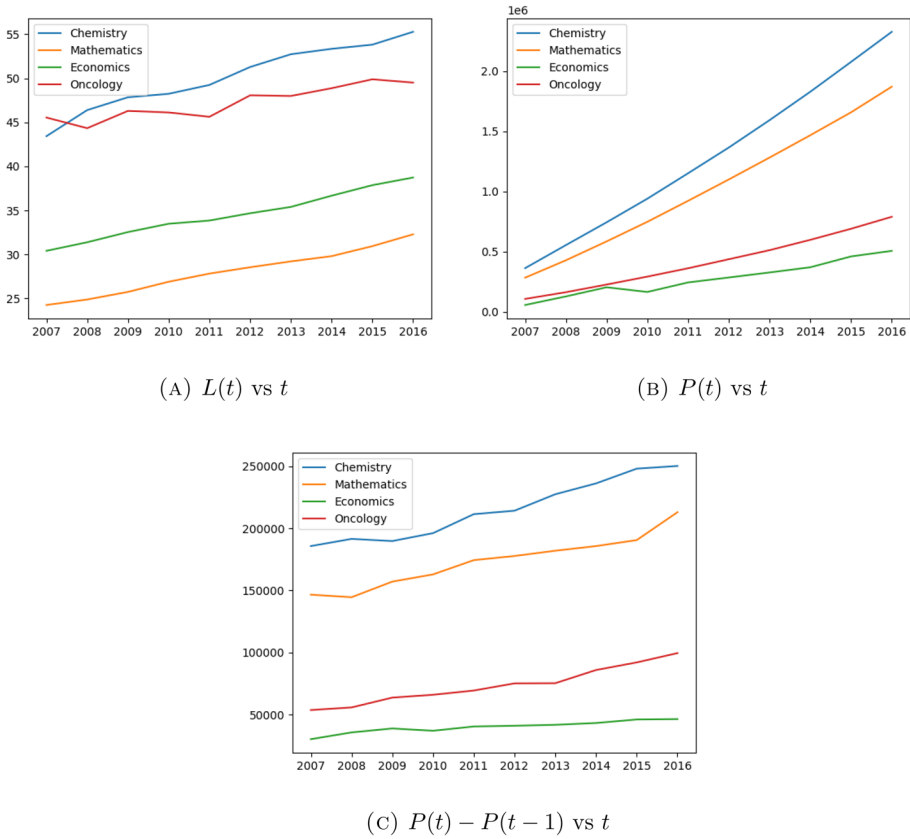(B) $P(t)$ vs $t$



(C) $P(t) - P(t-1)$ vs $t$

**Fig. 1** Relationships between $t$, $P(t)$ and $L(t)$

## Our model and the exponential growth phase

Our uniform model predicts an exponential growth rate of reference list length during a period in which growth in articles is exponential. The data supplied by the authors of Lariviere et al. (2007) contradicts this clearly: the growth of reference lists, the growth in $P$, and the relationship between $L$ and $P$ in various aggregated subfields over the period 1900–2004 are shown in Fig. 4.

In fact no function $q$ in our more general model will yield a subexponential growth rate of reference lists when literature is growing exponentially, as the following computation shows, assuming $P(t) = Ce^{kt}$:
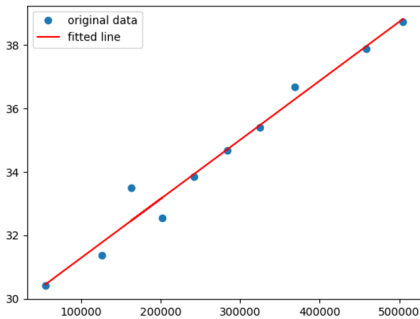
$$L(t) = Ck \int_{-\infty}^{t} q(t-s)e^{ks} \, ds$$
$$= Ck \int_{0}^{\infty} q(a)e^{k(t-a)} \, da$$
$$= Cke^{kt} \int_{0}^{\infty} q(a)e^{-ka} \, da.$$
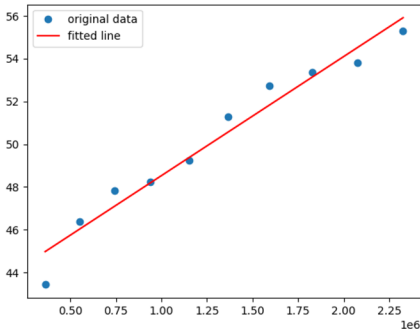
(A) Chemistry – Slope: $6 \times 10^{-6}$
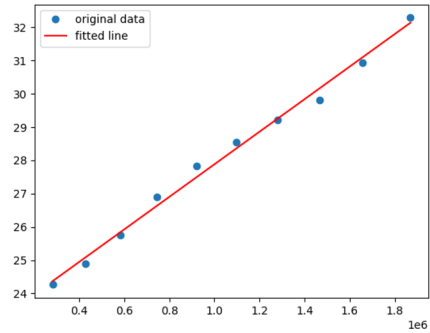
(B) Mathematics – Slope: $5 \times 10^{-6}$

(C) Economics – Slope: $19 \times 10^{-6}$
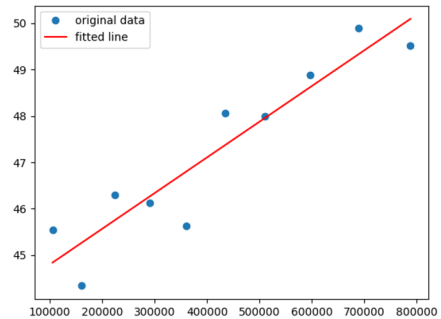
(D) Oncology – Slope: $8 \times 10^{-6}$

**Fig. 2** Linear relationship between $P(t)$ and $L(t)$
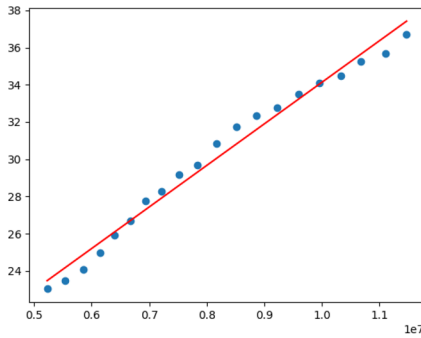
The integral in the last line is a nonnegative constant.

These results are not surprising, since the justification for the model is lacking in this case. It is not reasonable to assume that during the exponential growth phase of science (before the 1980s), authors could consistently locate all relevant references. For example, that time period was before the advent of widespread electronic search for publications. A more reasonable assumption is that authors added a few references to the reference lists of articles that they had read. This would lead to fairly slow growth in $L$, perhaps a little faster than linear.

## Connection to the literature on obsolescence and aging
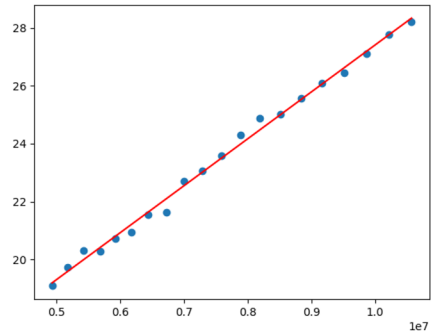
### Obsolescence and aging

A seemingly very different topic in bibliometrics is the aging/obsolescence of scientific literature. Having reviewed at least 30 articles from the bibliometric literature on these topics, we make several critical comments about that literature.
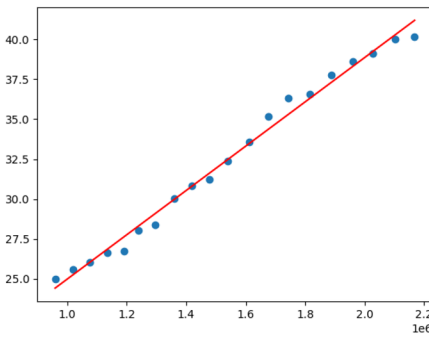
First, we deal with terminology. Many authors use "aging" as a synonym for "obsolescence". For a diachronous example, Brookes (1972): "Aging of scientific and
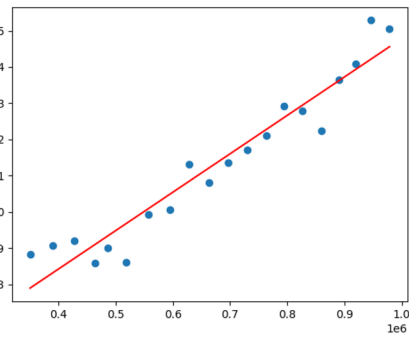
**Fig. 3** Linear relationship between $P(t)$ and $L(t)$ from dataset of Lariviere et al. (2007), 1985–2004

**Table 2** Fit of linear model for $L(t)$ versus $P(t)$ from dataset of Lariviere et al. (2007), 1985–2004

| Field | $R^2$ |
|---|---|
| Medicine | 0.982209 |
| Natural sciences | 0.994795 |
| Social sciences | 0.990064 |
| Arts and humanities | 0.925461 |

technical literature in this case means a decrease in intensity (frequency) of the use of publications by scientists and specialists with the increase in the time which elapsed since the date of publication. To measure the obsolescence rate, ...”; A synchronous example is Egghe & Ravichandra Rao Egghe (1992): “The paper deals with the shape of the obsolescence function, which one can construct, based on the age data of reference lists.” (Gupta, 1998), p. 336 states clearly as justification for a synchronous approach to obsolescence that “Evidence of the obsolescence of publications is presumed, if the use of these publications declines with age. Decline in the use literature over time (aging or
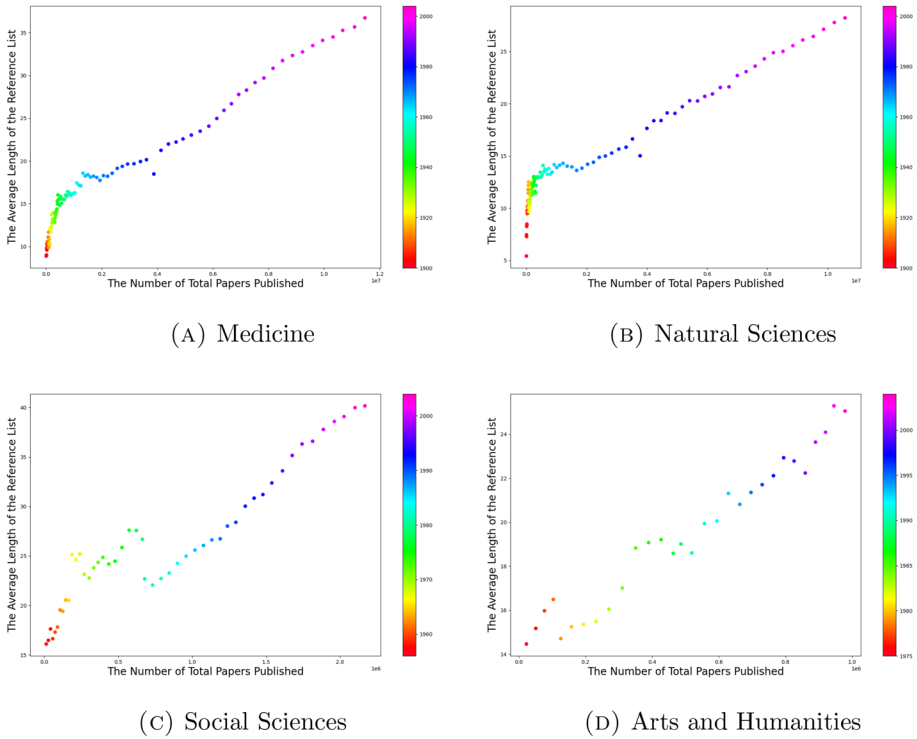
(A) Medicine



(B) Natural Sciences



(C) Social Sciences



(D) Arts and Humanities

**Fig. 4** Relationship between $P(t)$ and $L(t)$ from dataset of Lariviere et al. (2007), 1900–2004

decay) may be ascertained through studies of library use or by studying age of citations in publications or articles cited."

The above examples conflate two different concepts. Obsolescence means a lowered utility (in our context, being less attractive for citation) occurring over time. However, as we see below, obsolescence may occur as the article gets older, but it is not very accurate to say that an article becomes obsolete just because it is older. Rather, it is also because more relevant, more citable articles have been published in the interim. By analogy, steamships did not become obsolete simply because time passed, but because better alternatives were invented.

Also, some authors have used the term "aging" to refer to an increase over time in the (mean or median) age of items occurring in reference lists—in other words, aging (en masse, not individually) of reference lists. This usage may easily lead to confusion, because increase in the age of reference lists is caused by a *decrease* in the rate of "aging" (in the sense of obsolescence) of articles!

To avoid confusion, from now on we refer only to obsolescence of articles and recency bias of reference lists.

There is an even more important point to make—obsolescence is often confused with growth of literature, and is not really proven. The connection of our work on reference list length to these topics is as follows.

(1)   Growth in scientific literature leads to recency bias in reference lists: older items are crowded out by the more numerous newer ones. This is true even if the length of reference lists keeps pace with the growth in article production. It is not necessary to postulate any obsolescence effect (the decline in citability of articles as time passes) in order for recency bias in reference lists to occur.

(2)   However, any age-related obsolescence effect will also cause recency bias, even if no growth in the literature occurs. Thus, recency bias in reference lists is caused by both growth of literature and obsolescence.

(3)   Many authors have taken obsolescence (that is, Hypothesis B) as given, and used recency bias in reference lists as a way to estimate obsolescence.

The first point holds because this behavior is seen in our uniform model, as discussed in "The uniform model and the age distribution" below. In this model, the fraction of papers cited at time $t$ and having age at least $a$ is simply $P(t-a)/P(t)$, which decreases as $a$ increases. This obvious effect has been noted before. For example, in Lariviere et al. (2007) we find: "we suggest that the principal cause of the increased age of the cited scientific information is a fairly mechanistic response to the phenomenal growth in the quantity of published material after the war and to the current slowing in the growth of science."

The second point is clear. If $q$ is a decreasing function of $a$, at least after a short initial period, then we would expect this expression for $L(t)$ to be smaller than when $q$ is constant.

We come now to the third point listed above. Of course, if obsolescence is conflated with aging, it is obvious, because articles do age! Beyond that, Line and Sandison (1974) point out that obsolescence of articles is an assumption that must be verified, and there is evidence that it does happen. Many other authors have simply assumed obsolescence of articles. For example, Lariviere et al. (2007) assume obsolescence (after, as usual with such an assumption, an initial increase in citability) as common knowledge. After a diasynchronous study of the age of items in reference lists of articles published in the period 1900–2004, they find: "Indeed, in contrast to a widely-held belief, scientific literature does not become obsolete faster nowadays and, actually, quite the opposite is observed. The useful life of scientific publications has been increasing steadily since the mid-seventies." This is correct insofar as an increase in the rate of obsolescence should lead to a reduction in the average age of items in reference lists, all other things being equal. However, the existence of obsolescence has not been shown by this study—the results on recency bias in reference lists can be explained without it, as explained above.

Line and Sandison (1974) surveyed the area almost 50 years ago, and most of our observations were already noted by them or earlier authors. However, as described above, the distinction between obsolescence of articles and recency bias in reference lists, and the fact that both obsolescence and growth of the literature play a role in recency bias of reference lists, has subsequently been missed by several authors.

Some of the confusion between obsolescence and recency bias may be caused by a mistake in conditional probability: we want to know whether $P(cited \mid old)$ is small, but we only observe $P(old \mid cited)$. Bayes' formula shows

$$P(cited \mid old)P(old) = P(old \mid cited)P(cited).$$

Since $P(old)$ may be considerably smaller than $P(cited)$, we cannot make the desired conclusion.

In Lariviere et al. (2007) we find: "we suggest that the principal cause of the increased age of the cited scientific information is a fairly mechanistic response to the phenomenal growth in the quantity of published material after the war and to the current slowing in the growth of science." However, growth by itself cannot explain recency bias in reference lists, as our uniform model shows. A consequence of our model is that a slowdown in article production will lead to an increase in the average age of items in reference lists. However, if obsolescence is really a factor, then this property need not hold.

In summary, the relationship between aging and obsolescence is considerably more nuanced than much of the literature would lead a casual reader to believe.

As a more general criticism, we claim that the literature on obsolescence is of rather limited practical relevance today, and some of it appears to have been misconceived from the start. As explained by Line and Sandison (1974), we should distinguish between obsolescence of documents and of knowledge. In the age of print collections, limitations on shelf space made it important for librarians to determine which serials were less likely to be consulted in future by their patrons. However, in the electronic age where it is possible to maintain easy access to the long tail of less popular items, obsolescence of documents is much less of a concern. It is not clear whether obsolescence of knowledge is a major problem, or ever was.

In short, we find the literature on aging/obsolescence to be in serious need of improvement, and hope that our contribution here will spur further more rigorous work.

## The uniform model and the age distribution

Consider, under our general model, the reference list of articles published at time $t$. The expected total age of all items in such a list is

$$A(t) := \int_{-\infty}^{t} (t - s)q(t - s)P'(s)\,ds$$

Note that $A(t)/L(t)$ gives the mean age of an item in the reference list of a randomly selected article published at time $t$.

We restrict now to the uniform model, where $q(a)$ is constant for all $a \geq 0$. Since it will be difficult to analyse this in full generality owing to lack of historical data on $P$, we restrict to citations to articles published at time $t_0$ or later by articles appearing at time $t_0$ or later, to obtain the formula

$$A^*(t) = q \int_{t_0}^{t} P^*(s)\,ds$$

Integration by parts allows us to rewrite this quantity as

$$A^*(t) = q \int_{t_0}^{t} (t - s)P'(s)\, ds$$

$$= qt \int_{t_0}^{t} P'(s)\, ds - q \int_{t_0}^{t} sP'(s)\, ds$$

$$= qt\big(P(t) - P(t_0)\big) - q[sP(s)]_{t_0}^{t} + q \int_{t_0}^{t} P(s)\, ds$$

$$= q \int_{t_0}^{t} P(s) - q(t - t_0)P(t_0)\, ds$$

$$= q \int_{t_0}^{t} \big(P^*(s) - P(t_0)\big)\, ds$$

$$= q \int_{t_0}^{t} P^*(s)\, ds$$

We use a "mean-field approximation" approach, estimating the expected mean age of an item by

$$\frac{A^*(t)}{L^*(t)} = \frac{\int_{t_0}^{t} P^*(s)}{P^*(t)} \tag{1}$$

and the median by solving $P^*(t - a) = P^*(t)/2$ for $a$. Note that the expected mean age is is at most $t - t_0$, and equals the reciprocal of the fraction of the rectangle formed by the points $(t_0, P(t_0))$, $(t, P(t_0))$, $(t, P(t))$ and $(t_0, P(t))$ that lies under the graph $y = P(t)$. If for example $P^*$ grows quadratically in $t$ (corresponding to a linear increase in the number of articles produced per unit time), then these approximations yield a mean age of $(t - t_0)/3$ and median $(t - t_0)(1 - 2^{-1/3}) \approx 0.206(t - t_0)$. We tested Eq. (1) on our dataset—Table 3 shows the predictions. To obtain data on age of items in reference lists, we randomly sampled 30 articles from each field for publication year 2016, and extracted their data from CrossRef. This required manual queries: because the details about a publication's references are not available using CrossRef's API, the entire reference list was manually extracted for each item. Afterward, a Python program was used to analyse the reference list and calculate the sample mean and median of the referenced items. Only the references published after 2006 were considered. The results are also shown in Table 3. The fit is good given the sample sizes.

It is not to be expected that our mean-field approach should lead to accurate predictions about the *distribution* of the length of the reference list for a randomly chosen article published at time $t$. The model predicts that this should follow a binomial distribution $Bin(P(t), c)$.

Table 3 Predictions versus sample statistics for age of references conditional on article being published in 2016 and references being published no earlier than 2006

| Field | Mean age (prediction) | Mean age (sampling) | Median age (prediction) | Median age (sampling) |
|---|---|---|---|---|
| Chemistry | 5.36 | 4.74 | 5 | 5 |
| Mathematics | 5.40 | 5.19 | 5 | 5 |
| Economics | 4.62 | 5.37 | 5 | 5 |
| Oncology | 5.65 | 4.79 | 5 | 5 |

However we do consider this more stringent test of the model in Fig. 5. The results are not especially good, as expected, and more detailed modeling of the distribution of reference list length should be undertaken.

## Discussion and conclusion

Our results show that the uniform citation model we have used is good enough to explain the time evolution of the length of reference lists and to relate this to the growth in the number of articles published, at least in the modern era of electronic searching. We conclude that the model is promising for statistical studies of this type, and may be of use in other areas. Our more general model may yield more detailed predictions, but we face the serious problem of how to estimate the obsolescence function $q$. It would be very interesting to study a fast-growing modern field which essentially started during the last 20 years, to see how well our uniform model holds up.

Differences in citation patterns of research fields were already discussed by Price (1965). Fields such as mathematics seem on the face of it to be more suited to the uniform model, where citability of results, once established, presumably does not decline as quickly as in more fast-moving fields. We hope to inspire other researchers to investigate citation
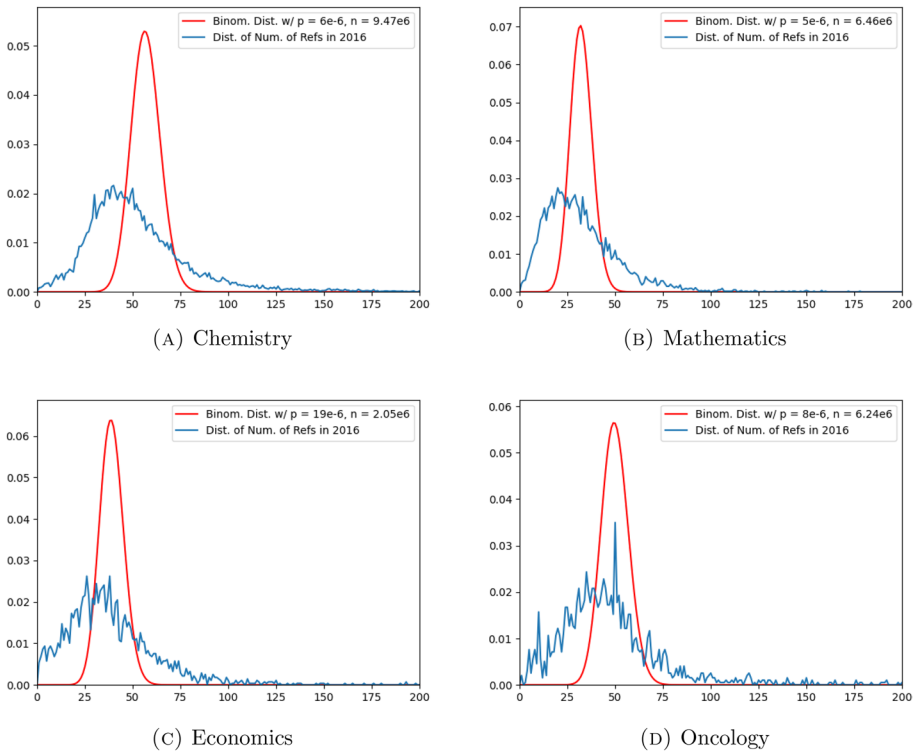


(A) Chemistry

(B) Mathematics

(C) Economics

(D) Oncology

**Fig. 5** Binomial fitting of mean reference list length of documents published in 2016

behavior in enough detail to allow improvements on the model presented here, and to distinguish more between growth and obsolescence.

The empirical and modeling results above lead to policy questions. The increasing growth in the number of references per article (also called "citation inflation") is potentially a cause for concern, if it continues long enough. This has been noted by previous authors (Ucar et al., 2014; Yitzhaki & Ben-Tamar, 1991). Given a constant cost per unit time for reading, scholars will face increasing difficulties in checking references. For example, our linear models predict an increase in the mean of at least five references per decade, and this would likely increase. This points to the need for more efficient machine reading of articles, which in turn requires open access to primary research literature. Another, more radical, option would be for journals to restrict the number of references. This is analogous to applicants for grants or promotions focusing on the most important publications. Of course this would not reduce the work of the author, but it would help the reader, and make it easier to compare citation productivity of researchers over time. Alternatively, reasons for citation could be clarified and references to literature standardized. For example, all perfunctory (Moravcsik & Murugesan, 1975) citations (where the author mentions work by others in order to convince reviewers that the author is knowledgeable or to show respect to other researchers, but that work is not directly built on by the present paper) could be listed in a standard place where readers could classify them as such, and ignore them if time is short. This might be a nontrivial saving; for example, Cano (1989) found in a small study that over one-quarter of citations fall into this category.

## Appendix A. Datasets

The following description of the dataset was supplied by Academic Analytics.

We culled the Digital Object Identifiers (DOIs) of journal articles (co)authored by scholars in the Academic Analytics commercial database (AAD; http://www.academicanalytics.com/) between 2007 and 2019. The AAD is composed of an annually updated roster of faculty members employed by 390 American Ph.D. granting universities. Faculty members in the AAD are linked to each CrossRef-DOI journal articles they (co)authored using manual and semi-automated disambiguation. The AAD also contains the academic department affiliation(s) of each faculty member; these departments, in turn, are manually classified into disciplines based on NCES CIP codes (https://nces.ed.gov/ipeds/cipcode/). We extracted from the AAD the DOI of each journal article (co)authored by scholars whose departments are classified in the following disciplines: Chemistry, Economics, English Language and Literature, Geography, History, Mathematics, and Oncology. Each DOI was entered as a search term in the CrossRef API (https://www.crossref.org/) to retrieve the length of the reference list and the first and last page numbers, from which the total page count of each article was calculated.
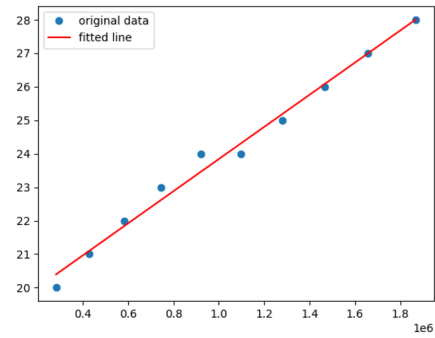
## Appendix B. Robustness checks

### B.1. Median vs. mean

We repeated the analyses using the mean from the main text, this time using the median. The results were similar, as seen in Figs. 6 and 7 and Table 4).
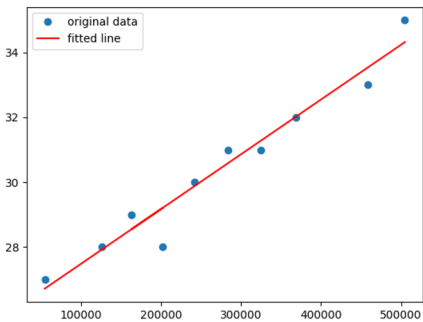
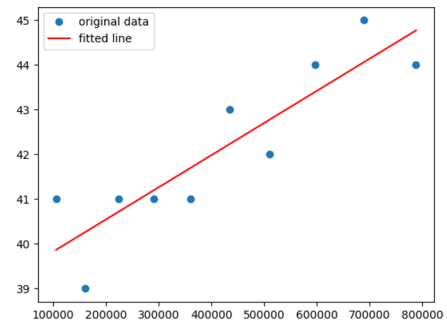**Fig. 6** Median reference list length, by year
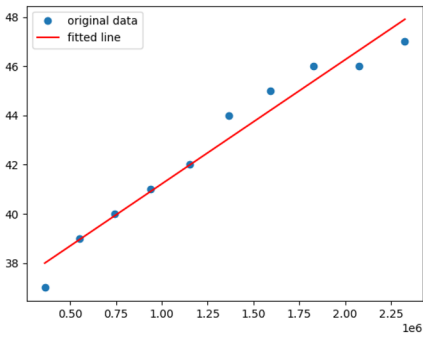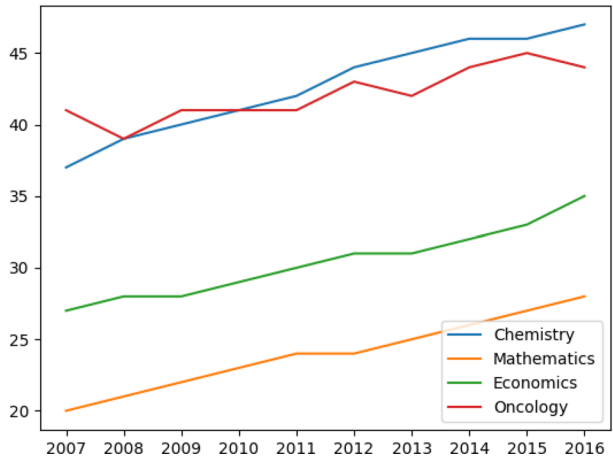




(A) Chemistry – Slope: $5 \times 10^{-6}$



(B) Mathematics – Slope: $5 \times 10^{-6}$



(C) Economics – Slope: $17 \times 10^{-6}$



(D) Oncology – Slope: $7 \times 10^{-6}$

**Fig. 7** Linear relationship between $P(t)$ and $L(t)$ (Median)

**Table 4** Fit of linear model for $L(t)$ versus $P(t)$ - Median

| Field | $R^2$ |
|---|---|
| Chemistry | 0.962511 |
| Mathematics | 0.988723 |
| Economics | 0.980056 |
| Oncology | 0.805837 |

## Declarations

## References

Boyack, K. W., et al. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. *Journal of Informetrics, 12*(1), 59–73.

Brookes, B. C. (1972). The aging of scientific literature. In A. I. Cherny (Ed.), *Problems of Information Science: collection of papers.* International Federation of Documentation, Study Committee, Research on Theoretical Basis of Information.

Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science, 40*(4), 284–290.

Egghe, L. (1992). Citation age data and the obsolescence function: Fits and explanations. *Information Processing and Management 28.2* (pp. 201–217). Elsevier.

Ellis, D. (1993). Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly, 63*(4), 469–486.

Gupta, B. (1998). Growth and obsolescence of literature in theoretical population genetics. *Scientometrics, 42*(3), 335–347.

Larivière, V., Archambault, É., & Gingras, Y. (2007). Long-term patterns in the aging of the scientific literature, 1900–2004. *Proceedings of the 11th conference of the International Society for Scientometrics and Informetrics (ISSI)* (pp. 449–456). ISSI.

Line, M. B., & Sandison, A. (1974). Progress in documentation: 'Obsolescence' and changes in the use of literature with time. *Journal of Documentation, 30*(3), 283–350.

Marx, W., & Bornmann, L. (2016). Change of perspective: bibliometrics from the point of view of cited references—a literature overview on approaches to the evaluation of cited references in bibliometrics. *Scientometrics, 109*(2), 1397–1415.

Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science, 5*(1), 86–92.

Price, D. J. D. S. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science, 149*(3683), 510–515.

Sugimoto, C. R., & Vincent, L. (2018). *Measuring research: What everyone needs to know*. Oxford University Press.

Ucar, I., et al. (2014). Growth in the number of references in engineering journal papers during the 1972–2013 period. *Scientometrics, 98*(3), 1855–1864.

Yitzhaki, M., & Ben-Tamar, D. (1991). Number of references in biochemistry and other fields. A case study of the journal of biological chemistry throughout 1910–1985. *Scientometrics, 21*(1), 3–22.