Noname manuscript No.
(will be inserted by the editor)

Marek Szydłowski · Adam Krawiec

Growth cycles of knowledge

Received: date / Accepted: date

Abstract We have developed a way of describing the increase with time of the number of papers in a scientific field and apply it to a data base of about 2000 papers on symbolic logic published between 1666 and 1934. We find (a) a general exponential increase in the cumulative total number of papers, (b) oscillations around this due to the appearance of new ideas in the field and the time required for their full incorporation, and (c) exogenously caused fluctuations due to wars and other non-scientific events.

1 Introduction

One of the most interesting aspects of a development of societies and civilisations is a role of knowledge understood as accumulated facts and experiences. In this way the knowledge can be naturally interpreted as science and technology. Therefore, we investigate the evolutional models of knowledge derived

The paper was supported by the Marie Curie Actions Transfer of Knowledge project COCOS (contract MTKD-CT-2004-517186). We would like to thank the anonymous referee for very interesting comments and suggestions. We are also very grateful to prof. Z. Roskal for discussion. We would like to thank dr Paweł Polak for help with preparing the data set.

Marek Szydłowski
Department of Theoretical Physics, Catholic University of Lublin, Al. Racławickie 14, 20-950 Lublin, Poland
E-mail: uoszydlo@cyf-kr.edu.pl

Adam Krawiec
Institute of Economics and Management, Jagiellonian University
Gronostajowa 3, 30-387 Kraków, Poland
E-mail: uukrawie@cyf-kr.edu.pl

Marek Szydłowski · Adam Krawiec
Complex Systems Research Centre, Jagiellonian University
Reymonta 4, 31-059 Kraków, Poland

from the scientometrics analysis of science, trying to describe its increase with time.

There have been many propositions on the mechanisms of science development so far. De Solla Price introduced the hypothesis of an exponential growth of scientific results and verified it on a large historical data set [4]. His research opened discussions in the scientometrics community which inspired many alternative roads to tackle the problem. As a consequence the statistical investigation of science became popular [8].

Another approach is based on an analogy with ecological models of growing population (e.g., [12]). Puzikov and Kasjanov [13] analysed the expansion of money spending on science in the USA using the Lotka-Volterra equation. In Goffman's approach, the processes of the development of a scientific field is treated as isomorphic with the processes of epidemic spread [5].

Let us notice that knowledge plays a very important role in the economic growth theory. Here, knowledge is treated as an important production factor which is one of the main causes of wealth in modern capitalistic societies. It is usually assumed that knowledge grows with a constant rate and it is treated as an exogenous factor. There are also many attempts at parametrisation of knowledge in an endogenous way, e.g. as human capital [14, 1].

As is well known, various economic and social processes can successfully be modelled in terms of systems of differential equations called dynamical systems. Since the process of growing science is similar to other economic and social processes (it is, in fact, itself a social process) as far as the degree of its complexity is concerned, treating it as a dynamical system seems naturally.

We have chosen published papers and monographs in a symbolic logic as an indicator of growth of knowledge because there is a good data base of papers published in the field between 1666 and 1934 compiled by Church [2,3]. Moreover, some papers are categorised as important and very important because of their increasing impact on the development of the field. The expertise of Church in this subject justifies the value of this compilation of empirical material.

In the present work, we construct a model of scientific growth in a given domain. We prefer the term "growth" rather than a term "progress" not to presuppose, in the terminology itself, that this process has really a progressive character. It could happen that a given chain of works will finally lead to a blind alley, but in the meantime it could exhibit a growing character to which our model could apply. The assumptions on which our model is based are the following.

1. The growth in question is measured by the number of important publications appearing at a given time. When testing the model the importance of a scientific publication should be determined in an operational sense, for instance, on the basis of the Citation Index or by the decree of a scientific authority, as it was the case when Alonso Church composed a list of publications in logic for the period 1847-1932 and marked on it those items which he considered fundamental. We assume that the number of important publications determines the state of science at a given time. We use the term "state of science", but we mean the state of a given branch of science, or a research program, or a given theory, depending on the modelled domain.

2. The growth of science has a continuous character but there exists a finite time period T needed to build up a result of the fundamental character. Further we assume that T is constant. Of course, it would be more realistic to assume a certain probability distribution for T, but we postpone this for another occasion. Our assumptions concerning T are better satisfied in some branches of strict sciences, such as mathematics or mathematical physics than in those scientific domains that strongly depend on empirical results. Such a result could cause a serious discontinuity in the process of growth.

3. We also neglect the interactions between various scientific domains, and assume that the modelled domain is an isolated system that is modelled by the equation

$$\frac{dx}{dt} = \alpha x(t - T) \tag{1}$$

where α is a proportionality constant. The initial condition $y(t) = \phi(t)$ is defined on the interval $[-T, 0]$. This equation is the basis of the present work. We shall also consider its generalisations or modifications.

Equation (1) states that the rate of scientific growth is proportional to the number of important publications at the present time t minus the time period T required to built up a fundamental result. It is very well known from the history of science that a certain time must elapse between the first appearance of a given scientific concept or method and its full exploitation in science.

2 Model of growth of scientific results

The process of scientific growth is inseparably linked to the fact that some results are getting old and finally are replaced by new ones. The latter are usually more general than the former, and include them as "limiting cases". This process, as described by Müller [12], leads to the following equation

$$\frac{dx}{dt} = \alpha x(t) - \beta x^2(t). \tag{2}$$

Equation (2) can be solved for x and the initial condition $x(0) = 0$; the solution is

$$x(t) = \frac{\alpha}{\beta}\left(1 - e^{-\alpha t}\right). \tag{3}$$

This is the equation of the logistic curve. For $t = \infty$ it reaches the saturation $x_0(t) = \frac{\alpha}{\beta}$. Dynamical system (2) has a stable node at this point.

We can also discretise the model, i.e., assume $\frac{dx}{dt} = x_{t+1} - x_t$. In such a case, differential equation 2 assumes the form of a map

$$x_{t+1} = (\alpha + 1)x_t - \beta x_t^2, \qquad x_t \in \left[0, \frac{\alpha}{\beta}\right] \tag{4}$$

which is very well known in the literature because of its Feigenbaum universality and the presence of chaos. Let us notice that the variable x_t belongs to the interval $[0,1]$. We introduce the new variable $u_t = \frac{\beta}{\alpha+1} x_t$, $\beta \neq 0$ which gives us

$$u_{t+1} = r u_t (1 - u_t) \tag{5}$$

where $r = \alpha + 1$. Since $t \to \infty$, $x_\infty = \frac{\alpha}{\beta}$ then $u_\infty = \frac{\alpha}{\alpha+1} < 1$ which means that we have here the logistic map for $0 < u < 1$. It can be shown that, for $3 \leq r \leq 4$ equation (4) exhibits chaotic behaviour, whereas for $1 < r < 3$ there exist only "large" periodic orbits with no chaos.

We shall assume that β is a small parameter. If $\beta \cong 0$, the parameter α admits a natural interpretation as a constant rate of growth of the results. The parameter $1/\alpha$ has the dimension of time, it represents the time after which one has the e-fold growth of results or e-folding time (it is also called the Lapunov time).

In the scientometrics approach one is often interested in the double increase time defined as

$$\tau = \frac{\ln 2}{\alpha} = t_{\mathrm{Lap}} \ln 2 \cong \frac{2}{3\alpha} = \frac{2}{3} t_{\mathrm{Lap}}.$$

It follows from the analysis of Goffman and Harman [6] that one or two important works appear in 12.5-year cycles (in the period investigated by them). This gives us the double increase time equal to about 25 years. Today this time strongly depends on a specific nation, its richness, scientific domain, etc. If we assume that $\tau = 10$ years, we would have $\alpha = \frac{2}{30} \cong 0.07$.

The parameter α turns out to be correlated with the increasing rate of change of knowledge, considered as an exogenous constant in economic discussions. The rate of growing of knowledge is assumed to be 0.05 in the economic growth theory.

In this analysis we have assumed that the number of scientists is constant and normalised to unity and considered the cumulative total number of papers. In the next section we consider the model with varying population of scientists. We assume the simplest case of exponential growth of their number. It means that we will investigate the evolution of average number of papers per scientist. Note that we will study the average number of all papers without their importance.

3 Models with a delay time and constant rate growth of population of scientists

Dynamical system (1) is very simple because its right-hand side is linear. In general it can be postulated in the nonlinear form

$$\dot{x} = f(x(t-T), x(t))$$

where $f(t-T)$ is a homogeneous function of degree one. The simplest form of such a function is a linear function and the de Solla Price model is obtained when the delay is zero. Let us introduce the change of population of scientists as an important factor in the growth of scientific field. We assume that the population grows at the constant rate n

$$\frac{\dot{L}}{L} = n.$$

Let us introduce the variable

$$z = x/L$$

Then the model has the form

$$\frac{dz}{dt} = \alpha z(t-T) - nz(t).$$

This equation describes the evolution of the number of papers written by a scientist. We can see that if the rate of change of population growth is zero we obtain the de Solla Price model of exponential growth. The above equation is still linear but now the cyclic behaviour may appear because of the feedback between the delayed and non-delayed terms.

To show the existence of the cyclic behaviour in this model we use methods of the bifurcation theory. The role of a control parameter in the model is played by the delay parameter T. First, we consider the characteristic equation

$$\lambda = \alpha e^{-\lambda T} - n \tag{6}$$

Next we assume that the eigenvalue is complex $\lambda = \sigma + i\omega$. By inserting $\lambda = \sigma + i\omega$ we obtain

$$\sigma = \alpha e^{-\sigma T} \cos \omega T - n \tag{7a}$$

$$-\omega = \alpha e^{-\sigma T} \sin \omega T. \tag{7b}$$

The bifurcation value $T = T_{\mathrm{bi}}$ leading to periodic orbit is obtained by substituting $\sigma = 0$. Then, system (7) can be solved in an exact form

$$\omega_{\mathrm{bi}} T_{\mathrm{bi}} = -\arctan\left(\frac{\omega_{\mathrm{bi}}}{n}\right) + j\pi \tag{8a}$$

$$\omega_{\mathrm{bi}} = \pm\sqrt{\alpha^2 - n^2}. \tag{8b}$$

Let us notice that these equations are symmetric with respect to the reflection $\omega = -\omega$ which means that if λ is a solution then $\bar{\lambda}$ is a solution as well. Therefore, we may assume $\omega > 0$. From these equations we can find $\omega(T_{\mathrm{bi}})$, and then the period $P = \frac{2\pi}{|\omega(T_{\mathrm{bi}})|}$. The period of the principal cycle ($j = 0$) is given by

$$P \simeq \frac{2\pi}{\sqrt{\alpha^2 - \beta^2}}. \tag{9}$$

We can also show that the period of the principal cycle is distinguished, because it is the only period longer than the bifurcation parameter value. The interpretation of minor cycles with periods shorter then the bifurcation parameter value is unattainable. This bifurcation analysis yields the cyclic behaviour in the system for some T in the close neighbourhood of T_{bi}. In the phase space we obtain an infinite number of isolated closed orbits.

From the above solutions we can draw the following conclusions:
1. If $n = 0$ the period of the unique cycle that is not shorter than t_{bi} is

$$P = \frac{2\pi}{\alpha}.$$

2. We can readily find the dependence between the bifurcation time T_{bi} and the period of the principal cycle P. If $\alpha \gg n$ then

$$P = 4T_{\mathrm{bi}}.$$

3. From (8) we find that the bifurcation time T_{bi} (if $n = 0$) is given by

$$T_{\mathrm{bi}} = \frac{\pi}{2\alpha}.$$

In general

$$T_{\mathrm{bi}} = \frac{1}{\alpha\sqrt{1 - \left(\frac{n}{\alpha}\right)^2}} \arctan \sqrt{\left(\frac{\alpha}{n}\right)^2 - 1}. \tag{10}$$

The graph presenting T_{bi} as the function of α, for $n/\alpha = 1/3$ is shown in Fig. 1. As it can be observed, to generate a long-term cycle one must assume relatively large delay time; e.q., for $T_{\mathrm{bi}} = 25$ years (this is approximately the length of the active period in the life of a scientist), one obtains $P = 100$ years. The parameter n diminishes this value.

We can see that, for a value of T from the suitable neighbourhood of T_{bi}, the system has a periodic solution which in the phase space is represented by the isolated periodic orbit of a given period because the system considered is linear.

4 Dynamics of empirical sciences

When we consider only theoretical results in science we neglect one important factor which influences new ideas. Natural sciences have an empirical character and all observational and experimental data create their firm foundations. There is a feedback between theory and observations which we must take into consideration.

In this section we present analytical solutions of the boundary problem for dynamics of empirical sciences, where both theory and observations makes new results possible. On the one hand, following Blavais and Kochen we assume that new important papers are the result not only of rethinking the past knowledge but also analysing the data in a given domain of science. On the other hand the theoretical developments direct the empirical studies. Theory helps to project experiments indicating observables which we should measure to test hypotheses.

For example, many cosmological questions are expected to be answered when we get higher precision of measurements of fluctuations in the cosmic background radiation. This hope drives many satellite programmes in astrophysics. These missions increase the available data. There is evidence in

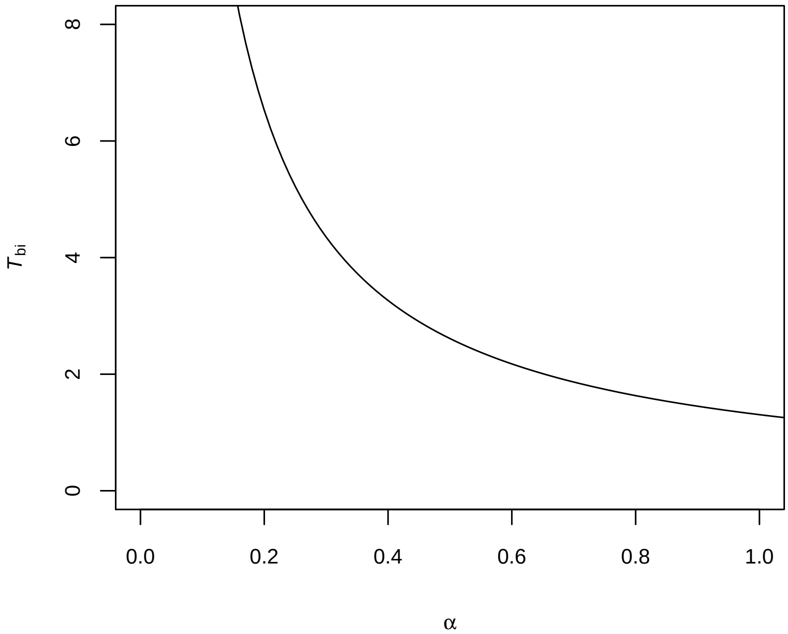

Fig. 1 The graph of T_{bi} with $n/\alpha = 1/3$ as the function of α given by (10).

different domains of science that the amount of data grows with a constant rate [15]. That is why we add it to our assumptions.

Let us formulate our model in terms of differential equations

$$\dot{x}(t) = \alpha x(t - T) + \beta \dot{y} \tag{11a}$$
$$\dot{y}(t) = gy(t) \tag{11b}$$

where x is a number of important papers and y is a measure of the empirical data. The number of new papers depends on the velocity of growing empirical results. More new data generate more new theoretical papers.

Equation (11) can be rewritten as a single non-autonomous equation

$$\dot{x} = \alpha x(t - T) + \beta g e^{gt}. \tag{12}$$

It is helpful to introduce a new variable z such that

$$z(t) = x(t)e^{-gt}. \tag{13}$$

Then

$$\dot{z} = \alpha e^{-gT} z(t - T) - gz(t) + \beta g \tag{14}$$

is an autonomous equation.

The constant term in (14) can be eliminated after moving a critical point of (14) z^* to the origin with the substitution

$$w(t) = z(t) - z^* \tag{15}$$

which gives

$$\dot{w} = \alpha e^{-gT} w(t-T) - gw(t) \tag{16}$$

where the critical point is

$$z^* = \frac{\beta g}{g - \alpha e^{-gT}}.$$

The further simplification can be obtained by another substitution

$$u(t) = e^{gt} w(t). \tag{17}$$

The final equation has the form

$$\dot{u}(t) = \alpha u(t-T). \tag{18}$$

This is a first-order delay differential equation which is our basic equation (1).

5 Analysis of bibliographic data

In the analysis of bibliographic data it is sufficient to consider a linear model with time delay (1) where the cumulative total number of papers is considered. It is interesting that it is possible to obtain the exact solution of (1) with $\alpha > 0$ in terms of eigenvalues of the linearisation matrix. The characteristic equation is of the transcendental form and admits the infinite number of complex solutions

$$x(t) = C_0 e^{rt} + \sum_{k=1}^{\infty} e^{p_k t}(C_{1k}\cos(q_k t) + C_{2k}\sin(q_k t))$$

where $\lambda_k = p_k \pm iq_k$ are complex roots of the characteristic equation $\lambda = \alpha e^{-\lambda T}$; r is a real root of the characteristic equation; and C_0, C_{1k}, C_{2k} are arbitrary constants. What is unexpected in this model is that it admits oscillations while the system is linear. The first term in the general solution represents the exponential part, while the sum describes fluctuations along this trend. We start with an estimation of this base trend.

Church prepared the bibliographic list of more than 2000 works in symbolic logic published in 1666-1934 [2,3]. The cumulative numbers of works in this period is presented in Fig. 2. One can observe that data exhibit an exponential trend with some periodic oscillation.

First we can estimate the exponential curve from this data. The standard analysis of linear regression for $\ln(x) = b + at$ where x is the cumulative number of works in time t, and b, a are constants, yields

$$\ln(x) = -40.4598 + 0.0243450t. \tag{19}$$

with $R^2 = 0.938232$.

The regression is shown in Fig. 3. We can see that some oscillations exist around on the exponential trend. In the previous section we argued that the

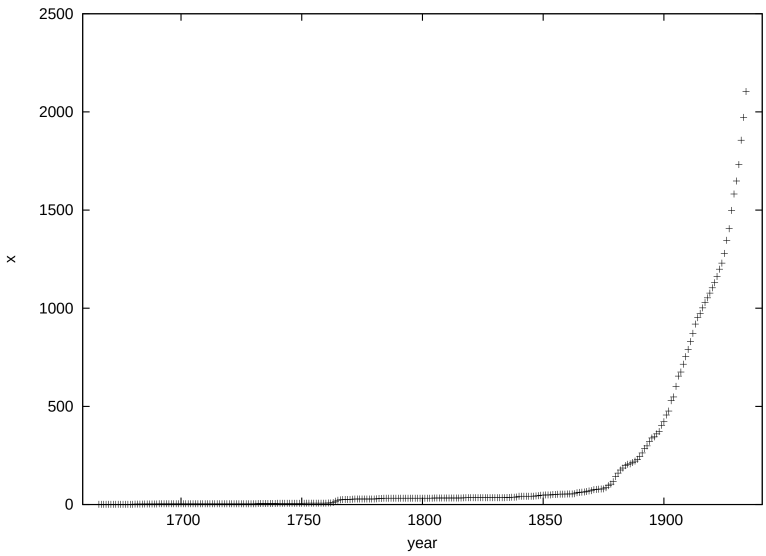

Fig. 2 The cumulative number of works x in period 1666–1934.

time required to build a result on the old ones is the cause of the oscillation around this exponential trend.

So far, we considered the delay in perception of scientific results as the main source of oscillations. On the other hand we cannot exclude the exogenous factors which have influence on science development. We cannot neglect the social and political events which should add to irregularities in the development of science. To corroborate this we use the CUSUM test for parameter stability [7]. The test is performed for regression of $\ln(x)$ with respect to year (the first 15 observations from the data set were excluded). The cumulated sum of the scaled forecast errors is calculated and presented in Fig. 4. The null hypothesis of parameter stability is rejected at the 5 percent significance level if the cumulated sum strays outside of the 95 percent confidence band. We can see that it happens in periods 1793-1798 and from 1917 onward. We can link up these periods with the French revolution and the World War I.

6 Conclusions

In this paper we presented a description of knowledge more realistic than the simple exponential growth. We showed that development of knowledge could be seen as an increase of the number of papers and monographs published in a given field of science. This growth has the exponential character but we found that there are some irregularities around the trend. As the cause of this fluctuations we considered the time delay interpreted as time passed between the publishing of a fundamental, important paper and papers based on it.

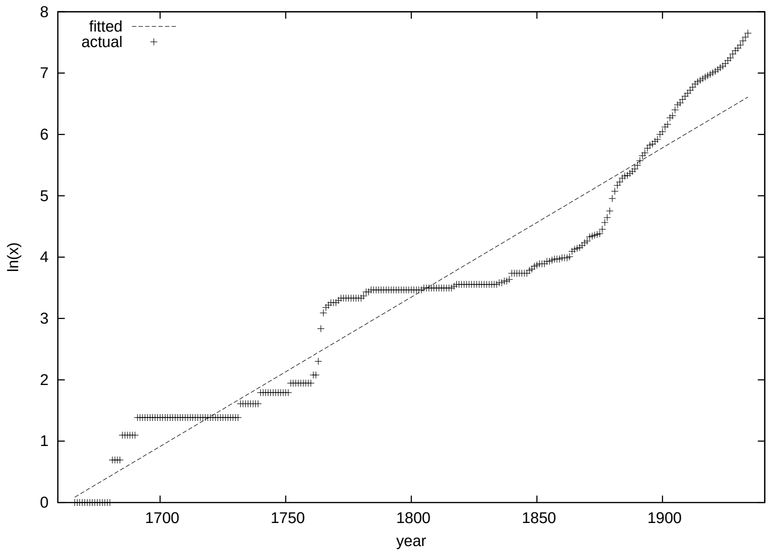

Fig. 3 The logarithm of cumulative number of works x and the regression line.

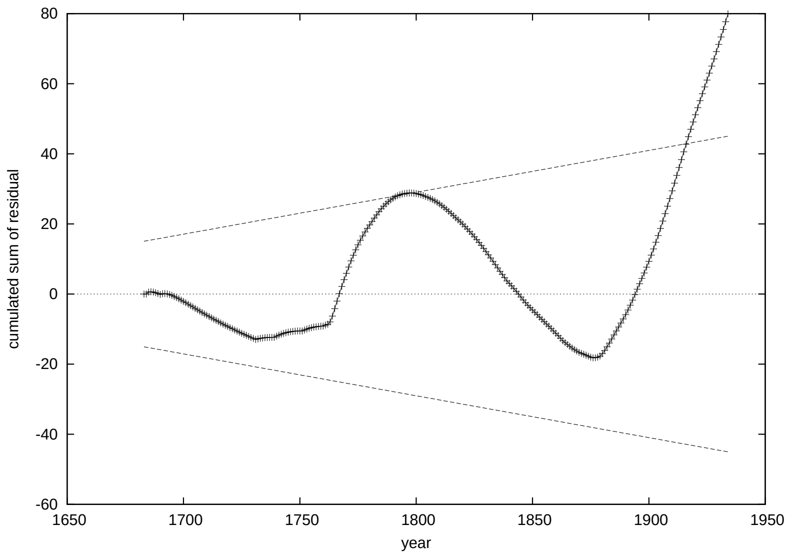

Fig. 4 The CUSUM test for parameter stability of regression (19) with 95% confidence level.

We have constructed a model of the growth of a field of science in which the growth is measured by the number of important works. We have also shown that in these models there exists a mechanism of a cyclic behaviour. This scenario corresponds to a bifurcation leading to periodic orbits. If the delay time (a time needed to built up a given scientific concept) is large enough, then the long term cycles appear with the period equal to about $4T_{\rm bi}$.

The above model has been generalised to the case when the population of scientists changes in a constant rate. We have demonstrated the existence of cycles in the number of important works per scientist.

The most general approach is to define a "production function" in science

$$Y = F(X)$$

where Y is a number of papers published in a given year, X the main factor which determine the amount of published papers (e.g., the number of all works in a given domain of science, spending on scientific research etc.).

Gupta and Karisiddappa [10] presented the model of development of cumulated number of papers. They argued that their model is more adequate than simple exponential and logistic models. In their model the rate of change of number of publications is of the power-law type

$$\frac{dx(t)}{dt} = at^b \tag{20}$$

where the constants a and b are positive. It is easy to see that their model can be reduced to the form

$$\frac{dx(t)}{dt} = \alpha x^{b/b+1} = \alpha f(x). \tag{21}$$

and then the "production function" $f(x)$ extracted from this model has the power-law form.

The great merit of these models is their simplicity. Their status could be compared with that of mechanical two-body models. Nobody has ever seen two isolated bodies in nature, but these models work well and are useful in explanations of many phenomena. Whether this is the case as far as our models are concerned, must be decided on empirical grounds by comparing the parameters appearing in these models with their values deduced from data referring to the real development of various domains of science.

References

1. R. J. Barro, X. Sala-i-Martin (2004) Economic Growth. Second Edition. MIT Press, Cambridge, Ma.
2. A J Church (1936) A bibliography of symbolic logic. Journal of Symbolic Logic 1:121–218
3. A J Church (1938) Additions and corrections to A bibliography of symbolic logic. Journal of Symbolic Logic 3:178–192
4. D J de Solla Price (1961) Science since Babylon. Yale University Press, New Haven
5. W Goffman (1966) Mathematical approach to the spread of scientific ideas — the history of mast cell research. Nature 212(5061):449–452
6. W Goffman, G Harman (1971) Mathematical approach to prediction of scientific discovery. Nature 229(5280):103–104
7. W H Greene (2000) Econometric Analysis, 4th edition, Prentice-Hall, Upper Saddle River, NJ

8. B C Griffith (1984) Derek Price (1922–1983) and the social study of science. Scientometrics 6:5–7
9. J Guckenheimer, P Holmes (1983) Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. Springer, New York.
10. B M Gupta, C R Karisiddappa (2000) Modelling of growth of literature in the area of theoretical population genetics. Scientometrics 49:321–355
11. M Kochen, A Blaivas (1981) A model for the growth of mathematical specialities. Scientometrics 3:265–274
12. F Müller (1972) Fortschritt der Wissenschaft – mathematical Modelliert. Wissenschaft und Fortschritt 22:162–165
13. M D Puzikov, A E Kasjanov (1987) Quantitative estimation of Big and Little Science interrelation. Scientometrics 11:99–104
14. D Romer (1996) Advanced Macroeconomics. McGraw-Hill, New York
15. A Szalay private information
16. M Szydlowski, A Krawiec (2001) Scientific cycle model with delay. Scientometrics 52:83–95