ON THE ANALYSIS OF LIBRARY GROWTH*

by H. L. Resnikoff and *J. L. Dolby*

1. *Introduction*

The single most striking statistical fact about library holdings is their rapid rate of growth. Whereas population growth in the United States proceeds at rates less than 1.5 percent, all major university libraries add at least 3 percent annually to their holdings (which consequently double every 23 years), and the Universities of Connecticut, Maryland, and Toronto are adding to their holdings at a 10 percent annual rate (doubling in less than 8 years) (Ref. 3). Such rapid growth to an ever-increasing extent determines the reaction to the collection of both library users and library management. It creates and sustains an ever-present pressure against the human, financial, and physical resources of the library, and limits opportunities for increasing access to the information buried in the growing archive.

In a previous study (Ref. 3) the nature and most obvious implications of library growth were examined. There it was concluded that for long periods, in some cases nearly 400 years, the current exponential pattern of holdings growth has sustained itself, subject only to local fluctuations representing the effect of major historical phenomena, but always returning to the steady certainty

*This paper presents the substance of a lecture delivered by the first named author for the rededication of the Fondren Library at Rice University in April 1969. Both authors are pleased to acknowledge the support of Office of Education Bureau of Research contract OEC-9-8-00292-0107, and to thank the publishers who have kindly given their permission for the reproduction of figures which first appeared in the indicated publications: Figure 3—reproduced from *Energy in the Future* by P. Putnam, by permission of Van Nostrand-Reinhold Company, a division of Litton Educational Publishing, Inc., Litton Industries (Princeton, New Jersey, 1953); Figures 6 and 7—reproduced from *The Biology of Population Growth* by Raymond Pearl, by permission of Alfred A. Knopf, Inc. (New York, 1925); and Figure 12—reproduced from *Science Since Babylon* by Derek J. de Solla Price, by permission of Yale University Press (New Haven, Connecticut, 1961).

of exponential growth. This important observation referred to the number of holdings arranged by date of *imprint*; it did not refer to growth in holdings arranged by date of *accession*. Although these two types of growth are related, they behave in significantly different ways. Even libraries that are much younger than the invention of printing show the exponential behavior of their collections as a function of imprint date, but they also usually show a much more complex behavior of holdings as a function of accession date. For example, the holdings of the Library of Congress as a function of accession year are displayed in Figure 1 (on semilogarithmic graph paper, see also Table I). It is evident from Figure 1 that there are several distinct periods of growth, each of which is approximately exponential, but with varying growth rates. In effect, the Library's growth rate was greatest when it was small; as it developed, a lesser rate of growth appeared. This pattern is easy to interpret, and occurs in many situations unrelated to library problems. Of more significance is the fact that this growth is essentially *piecewise exponential,* that is, it consists of consecutive growth periods, each of which is exponential (see Section 3).

In Figure 2 the growth of the Widener Library subcollection represented in Volume 7 (bibliography) of the Widener Shelf List (Ref. 15) is shown as a function of date of imprint. This distribution, which covers a period of more than 400 years, is exponential for almost 300 of these years with but minor deviations. It should be contrasted with Figure 1.

The significant conclusion which is suggested by these two figures, and confirmed by further studies, is that there are two kinds of library growth. Some care must be taken to distinguish them from each other. What we have called the *imprint* growth rate is related to the total amount and nature of all published materials; the other kind of growth is particular to the life cycle of each library: it can reasonably be called *accession* growth. It is obvious that accession growth must be influenced by the more fundamental growth of imprints but, except possibly in special situations that do not currently exist, the converse is not true. Accession growth does not influence imprint growth.

The management problems of a specific library will be composed of a complex combination of subproblems stemming from both types of growth (as well as from other variables). It is necessary to be able to separate these two kinds of subproblems to be able to analyze their relative importance and to be able to provide reliable projections of future requirements.

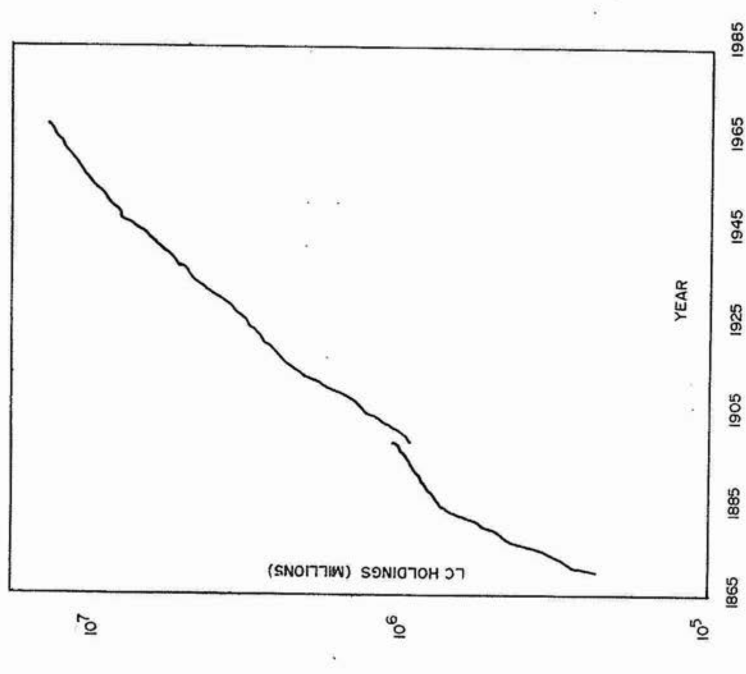


FIGURE 1. Library of Congress holdings, 1865–1966.

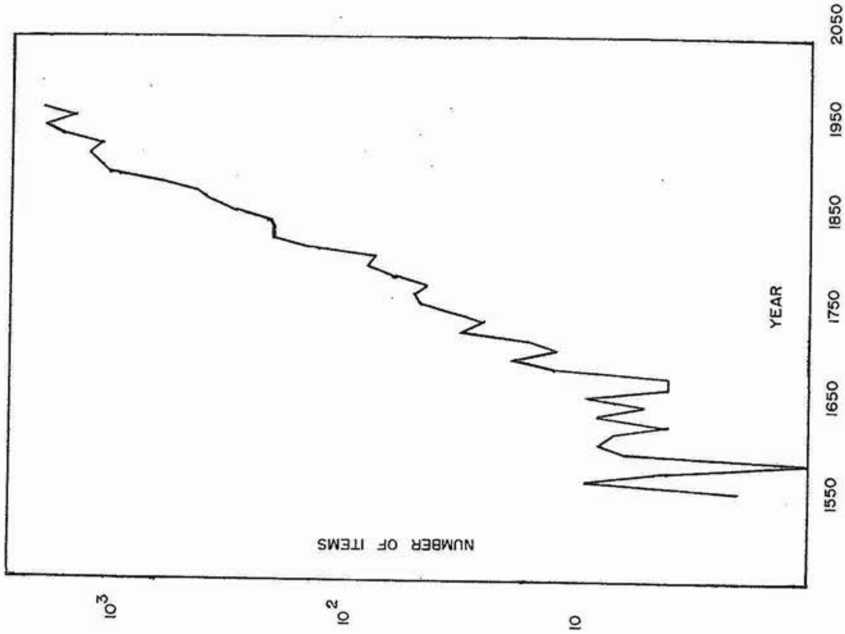


FIGURE 2. Imprint Date Distribution from the Widener Shelf List, Volume 7.

TABLE I

Library of Congress Holdings

Year	Pamphlets	Bound Volumes	Total Bound Volumes and Pamphlets	Bound Newspaper Volumes	Total Bound Volumes and Pamplets and Bound Newspaper Volumes
1866	~ 40,000	99,650			
1867		165.467			
1868		173,965			
1869		183,227			
1870	30,000	197,668	227,668		
1871	40,000	236,846	276,846		
1872	45,000	246,345	291,345		
1873	48,000	258,752	306,752		
1874	53,000	274,157	327,157		
1875	60,000	293,507	353,507		
1876	100,000	311,097	411,097		
1877	110,000	331,118	441,118		
1878	120,000	352,655	472,655		
1879	120,000	374,022	494,022		
1880	133,000	396,788	529,788		
1881	145,800	420,092	565,892		
1882	160,000	480,076	640,076		
1883	170,000	513,411	683,441		
1884	185,000	544,687	729,687		
1885	191,000	565,134	756,134		
1886	193,000	581,678	774,678		
1887	194,000	596,957	790,957		
1888	200,000	615,781	815,781		
1889	206,000	633,717	839,717		
1890	207,000	648,928	855,928		
1891	210,000	659,843	869,843		
1892	220,000	677,286	897,286		
1893	223,000	695,880	918,880		
1894	225,000	710,470	935,470		
1895	230,000	731,441	961,441		
1896	245,000	748,113	993,113		
1897	218,340	787,715	1,006,055		
1898	226,972	832,107	1,059,079		
			(932,094) Sic		
1899			957,056		
1900			995,166		
1901			1,071,647		
1902			1,114,111		
1903			1,195,531		
1904			1,275,667		
1905			1,344,618		
1906			1,379,244		
1907			1,433,848		
1908			1,535,008		
1909			1,702,685		
1910			1,793,158		
1911			1,891,729		
1912			2,012,393		
1913			2,128,245		
1914			2,253,309		
1915			2,363,873		
1916			2,451,974		
1917			2,537,922		

TABLE I (Continued)

Year	Pamphlets	Bound Volumes	Total Bound Volumes and Pamphlets	Bound Newspaper Volumes	Total Bound Volumes and Pamplets and Bound Newspaper Volumes
1918			2,614,523		
1919			2,710,556		
1920			2,831,333		
1921			2,918,256		
1922			3,000,408		
1923			3,089,341		
1924			3,179,114		
1925			3,285,765		
1926			3,420,345		
1927			3,566,767		
1928			3,726,502		
1929			3,907,304		
1930			4,103,936		
1931			4,292,288		
1932			4,477,431		
1933			4,633,576		
1934			4,805,646		
1935			4,992,510		
1936			5,220,794		
1937			5,395,044		
1938			5,591,710		
1939			5,828,126		
1940			6,102,259		
1941			6,349,157		
			6,353,516		
1942			6,609,387		
1943			6,822,448		
1944			7,304,181		
1945			7,877,002		
1946			7,946,460	118,159	8,064,619
			8,193,200		
1947			8,187,064	121,251	8,308,315
1948			8,387,385	124,619	8,512,004
1949			8,689,639	128,055	8,817,694
1950			8,936,993	131,425	9,068,418
1951			9,241,765	136,717	9,378,482
1952			9,578,701	140,573	9,719,274
1953			N/A		
1954			10,155,307	147,090	10,302,397
1955			10,513,048	151,623	10,664,671
1956			10,776,013	155,921	10,931,934
1957			11,037,773	159,015	11,196,788
1958			11,411,475	161,389	11,572,864
1959			11,779,894	165,741	11,945,635
1960			12,075,447	167,654	12,243,101
1961			12,329,678	169,993	12,449,671
1962			12,534,331	160,466	12,694,797
1963			12,752,792	156,766	12,909,558
1964			13,139,494	150,530	13,290,024
1965			13,453,168	149,509	13,602,677
1966			13,767,403	145,721	13,913,124

NOTE: The decline in Bound Newspaper Volumes 1961-1966 is undoubtedly due to the weeding out of old newspapers and replacement by micro-film.

It should not be thought that the observed long-term exponential growth of imprints will continue indefinitely. Physical resources, as well as author resources, cannot maintain the pace of exponential growth. It follows that the current phase of imprint growth must ultimately terminate, yielding to a type of growth, or perhaps decline, that will have an upper bound. It is important to know whether this time is come, or whether it still is far in the future, for in the first case we may breathe a sigh of relief, assured that our resources will come into compass with library requirements and perhaps even permit the luxury of elaborating access to library collections in a studied and leisurely manner not subjected to the current continual strain to catch up with the flood of accessions. In the second case, which the authors believe is the more likely of the two, exponential growth at least at current (and historic) rates will continue for a long period, and a relaxation of current accession pressures cannot be hoped for. In this case, every effort must be bent to command resources in the most efficient and rapid manner possible to cope with the information glut. It is in this connection that the computer must be considered, for the growth with time of the number of computer operations per second and the number of operations per dollar, are both exponential also and therefore provide the possibility of containing and assimilating the growing flow of information.

Because the current phase of exponential growth must ultimately change to some type of growth occurring at a lower rate, and ultimately to a growth or decline which has an absolute upper bound, it is of some interest to study the likely forms that future growth curves will take, as well as the way in which one growth stage will pass into the next. The remaining sections of this paper are devoted to determining the forms of natural growth curves that appear pertinent to the problem of library growth so that three main problems can be investigated. These are:

1. Is the current phase of exponential growth likely to persist for long?
2. What is the utility of approximating library and related growth curves by piecewise exponential functions in place of more complex curves?
3. What is the nature and interpretation of fluctuations about the exponential trend?

The answer to the first problem has the greatest immediate significance, for it determines whether future decades have in store for us exponential struggle, or the peaceful and perhaps dull coexistence with a steady state of information production and decay.

2. *Four Fundamental Communication Inventions*

The purpose of the library is to store and provide access to the information accumulated by man throughout history. Thus it is reasonable to expect a relationship between library growth and growth or change in other components of civilization. This fundamental relationship does not seem to have been sufficiently emphasized in the literature, but it is important for a balanced understanding of the current growth situation. Therefore, it seems appropriate to make a brief historical digression.

Four fundamental inventions can be recognized in the history of civilization. Each of them produced changes of unprecedented magnitude and extent; all of them have a common aspect. The first was the invention of complex writing systems, capable of expressing abstract concepts as well as names and actions, in the Near East not much earlier than 3000 B.C. (cf. Ref. 4). It was immediately followed by the rise of the oldest of the high civilizations known today—the Egyptian and Mesopotamian—and was perhaps the cause of their rise. These elaborate and inefficient writing systems made possible the accumulation of archival information stores containing records of complex processes and observations relating to mathematics, astronomy, law, political administration, and commercial accounts. There is no indication of the existence of complex political organizations or scientific effort in societies not having a writing system.

Nearly 2000 years later the alphabetic system of writing was introduced by the Greeks (Ref. 4),[1] and their remarkable civilization rose to its great heights shortly after. The significance of an alphabetic writing system cannot be overestimated: it is markedly more efficient for communication, requires less time to learn, and is therefore accessible to greater numbers of people. It provides a startling comparative advantage in communication for the maintenance of commercial and administrative records, for the recording of mathematical and scientific information, and even for those communications necessary for the extended prosecution of military efforts ranging over great distances. The mental effort and time required to read a given amount of information recorded in one of the ancient Egyptian writing forms or in cuneiform Akkadian is enormous compared with an alphabetic language representation.

The third invention that we consider fundamental is that of movable type and its application to printing, probably by Gutenberg and Johann Fust in the mid-fifteenth century; the earliest known book printed in Europe is dated 1456. Although movable type was known in China and Korea before its independent invention in Europe, its influence in China was negligible, no doubt because of the nonalphabetic nature of the Chinese language, while its development in Korea antedated the European invention by about 50 years, and was probably connected with the Korean adoption of a phonetic alphabet about that time.

Each of these three inventions provided a remarkable advantage over previous methods for *storing, transmitting,* and *retrieving* information, with a corresponding decrease in the unit costs involved. The principal consequence of increased efficiency and effectiveness was the opportunity to build on past knowledge and experience; this greatly accelerated the growth and progress of civilization in each of the periods of invention.

The modern general-purpose digital computing machine epitomizes the common properties of these three earlier inventions; no one other instrument or technique in the history of civilization has created such a change in the ability to store, transmit, and retrieve information as has the computer. It is difficult not to view it as the fourth fundamental invention in the field of communication of information, and to speculate that its effect will be no less than the effects of each of its predecessor inventions. From this standpoint, a new "information explosion" must be anticipated, that is, a new period of exponential growth having a growth rate (i.e., an exponent) greater than that characteristic of the post-printing press growth period. One consequence of this new explosion will be an *increased rate* of library accession of information and quite likely a new role for libraries as archives of information in machine-readable form. If this argument is accepted, the answer to the first of the questions posed in Section 1 is clear: the current exponential growth rate is *not* likely to persist; it will be replaced by an information growth rate *greater* than the current rate.

If the role of the computer in civilization is in fact similar to the roles previously played by the invention of writing systems, alphabet, and printing, then there will probably be changes in the essential fabric of civilization as we know it that cannot now be foreseen. The role of the library is central to this point of view, for libraries represent the archive wherein the knowledge and experience accumulated by previous generations is maintained and organized (albeit superficially); they are the information repositories on which all future developments are founded. If it is true that the growth of civilization, and indeed its growth *rate*, depends on the capabilities for storage, transmission, and retrieval of information, then it inevitably follows that the role of libraries as storage banks for information in printed as well as machine-readable form—and of computers to transmit and retrieve, as well as to analyze, modify, and re-store that information—must become more central and important as time passes. Governmental support for library systems, and library support for and experiments with computers is, if our argument is essentially correct, imperative for the continued growth of civilization. Examples of the interaction between efficient information-processing and the growth of civilization that may help to place the previous remarks in perspective are discussed below.

World Population Growth

The "population explosion" is a consequence of decreased death rates since there is relatively little that can be done to increase the per capita annual birth rate. As Figure 3 shows, the *rate of growth* of world population—that is, the decrease in the death rate—increased markedly after 1500, and has been increasing ever since in a historically unprecedented manner. This must be attributed primarily to the rapid communication of medical knowledge, sanitation techniques, advances in agriculture, and so forth, made possible by the invention of printing a few decades earlier. A similar population explosion probably occurred in the Greek world between 1000 B.C. and 500 B.C., and even earlier in historical Egypt and Mesopotamia, although accurate population estimates are lacking. The vast construction programs of the Egyptians and Akkadians, and the similar later ones of the Chinese, provide evidence of a massive long-term employment of labor and consequently of a relative surplus of population. It is of no importance to this argument what proportion of these populations might have been "immigrants" of one type or another.

Mathematical Research

Mathematical research depends in a direct way on the availability of an archive and the efficient transmission of information. Even in ancient times there was relatively rapid communication of important new results. For instance, Archimedes was well informed about the astronomical theories of Aristarchus although

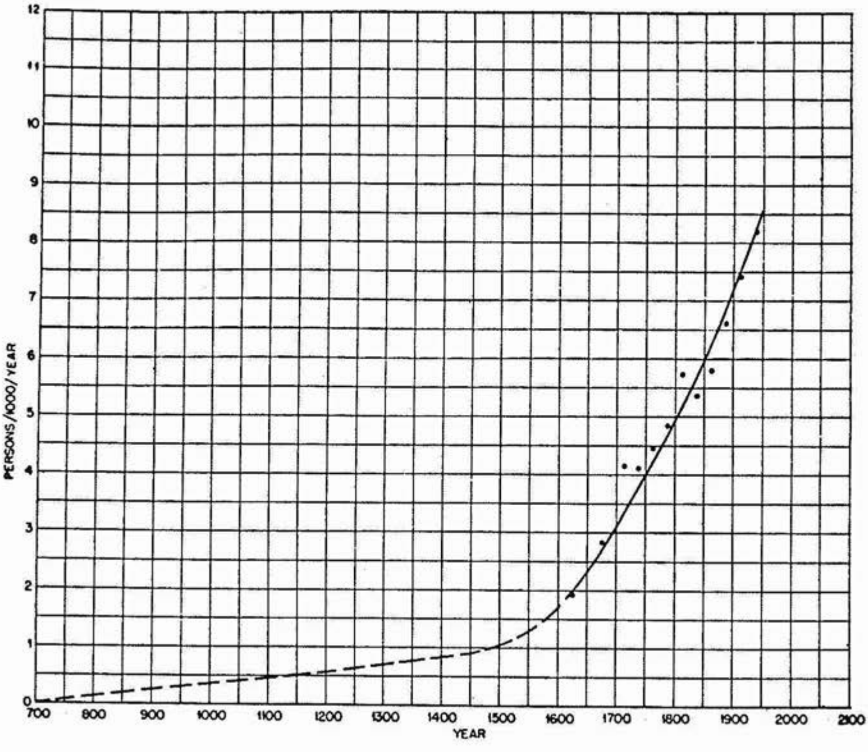


FIGURE 3. Estimated trend in the rate of growth of World population, A.D. 700 to 1950. *The point for 1951 was estimated by the U.N., 1952.* (Data from Palmer Cosslett Putnam, *Energy in the Future* [New York, 1953], p. 22.)

the latter was only 25 years older than Archimedes and they were separated by much of the "known" world of the times (circa 200 B.C.), Archimedes residing in Sicily and Aristarchus on the island of Samos near modern Turkey. Because of the critical role that communication plays in mathematical progress, one would expect that mathematics would experience a renaissance following an invention which decisively improves communication. In fact, there is no real mathematics known prior to the Egyptian and Mesopotamian civilizations, although serious mathematics of high caliber is attested in the earliest phases of these civilizations. Much more information is available about what has happened since Greek times, and it supports our view. Indeed, if the number of memorable mathematicians is graphed as a function of the birth date of the mathematician, then the curve displayed in Figure 4 results. Here a mathematician is "memorable" if he is named in the index to a standard history of mathematics; the book by Struik (Ref. 13) has been used for the illustration. Other choices would not change the shape of the curve. Estimates of world population are shown in Figure 4 for comparison purposes. Observe that the number of memorable mathematicians rose rapidly after 700 B.C.—that is, not long after the invention of an alphabet by the Greeks—and grew exponentially from 300 B.C. until about 1450 A.D. after which time the growth rate increased dramatically. If the exponential portion of the memorable mathematician curve is extended back in time, it suggests that the "first" memorable mathematician lived about 3700 B.C., not long before the first writing systems are attested and probably simultaneous with their development.

European Universities

The growth curve illustrating the currently extant European universities as a function of their date of founding (Figure 5) is interesting. The data came from the *Random House Dictionary*, and is given in Table II. Figure 5 shows four distinct phases of growth, three of which are clearly exponential, with a fourth that is approximately so. The earliest period, from 1100 to 1210, corresponds to such small numbers of universities that statistical arguments cannot be reliable, and we thus ignore it. The second period, from about 1210 to 1500, indicates a uniformly exponential growth doubling approximately every 110 years. After 1500, there is a 300-year period of roughly exponential growth, which may mark a transitional phase from the previous period to the next one, beginning in 1800 and continuing to the present, which shows

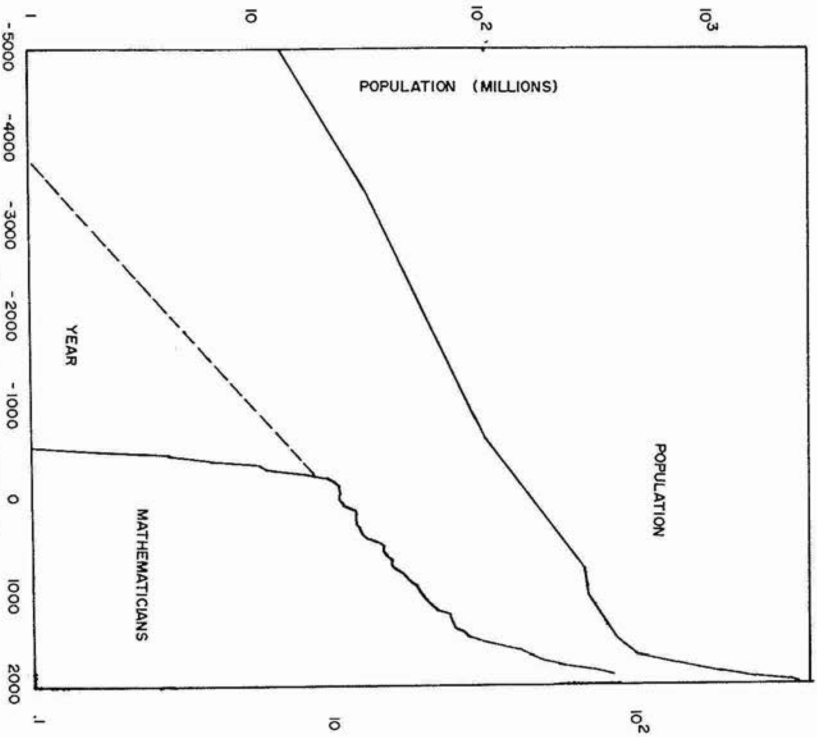


FIGURE 4. World Population and Memorable Mathematicians.

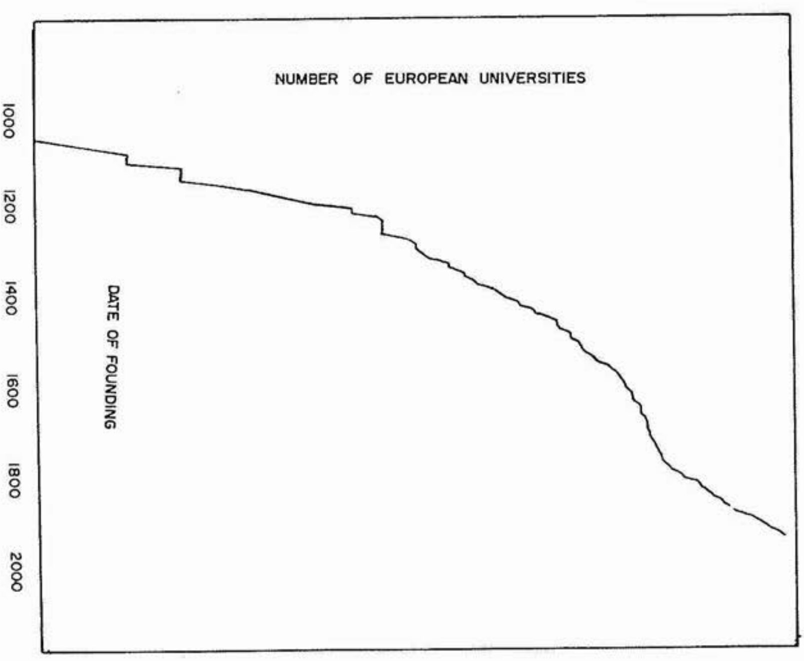


FIGURE 5. European Universities currently extant (includes Soviet Union).

exponential growth proceeding at the same rate *as the 1210 to 1500 growth.*[2] Thus shortly after the invention of printing a major change in the growth rate of the number of universities occurred. One interpretation of this phenomenon is that printing made the textbook cheaper and more available and thus permitted an increased student-to-teacher ratio, decreasing the necessity for founding new institutions. After 300 years, the natural growth in the student population increased this ratio beyond levels that could be efficiently maintained even with the availability of inexpensive texts, thus encouraging the foundation of more universities, at the previously observed rate. This explanation is offered solely as a possibility; research would have to be done to see if it is consistent with the growth of student populations. An indication that this explanation may not be farfetched is the contemporary effort to introduce television and teaching machines to permit greater student-to-teacher ratios; both of these devices are improved means for *transmitting* information, which of course the printed text is too.

3. *Stable Growth*

The growth with time of a population (of people, or books, or any other quantity) is said to be *exponential* if its rate of change is proportional to the population; the constant of proportionality is the *growth exponent*. In the standard notations of the differential calculus, this is expressed by writing

$$\text{1)} \qquad \frac{dP(t)}{dt} = a_o P(t) \ ,$$

where $P(t)$ denotes the population at time t, and a_0 is the growth exponent. The solution to this differential equation is

$$\text{2)} \qquad P(t) = P(t_o)\, e^{a_o (t-t_o)} \ ,$$

where t_0 is any conveniently chosen time. If a population does grow according to this equation, then its logarithm varies linearly with the time, that is,

$$\log P(t) = \log P(t_o) + a_o (t - t_o) \ ;$$

here log denotes the natural logarithm function. This means that exponential growth is represented by straight lines on semi-logarithmic graph paper, such as is used for Figures 1 and 2, for instance.

TABLE II

EUROPEAN UNIVERSITIES CURRENTLY EXTANT*

Year	Cumulative	Cumulative Minus Number in 1460 (=46)	Cumulative Minus Number in 1500 (=58)	Cumulative Minus Number in 1790 (=117)
1065	1			
1100	2			
10	2			
20	2			
30	3			
40	3			
50	3			
60	3			
70	4			
80	5			
90	6			
1200	7			
10	8			
20	11			
30	11			
40	13			
50	14			
60	14			
70	14			
80	14			
90	17			
1300	18			
10	18			
20	19			
30	20			
40	23			
50	23			
60	26			
70	26			
80	28			
90	29			
1400	32			
10	34			
20	36			
30	39			
40	40			
50	44			
60	46			
70	52	6		
80	52	6		
90	53	7		
1500	58	12	0	
10	58	12	0	
20	61	15	3	
30	63	17	5	
40	65	19	7	
50	68	22	10	
60	70	24	12	
70	75	29	17	
80	79	33	21	
90	82	36	24	

TABLE II (Continued)

Year	Cumulative	Cumulative Minus Number in 1460 (=46)	Cumulative Minus Number in 1500 (=58)	Cumulative Minus Number in 1790 (=117)
1600	85	39	27	
10	87	41	29	
20	88	42	30	
30	91	45	33	
40	92	46	34	
50	92	46	34	
60	97	51	39	
70	97	51	39	
80	98	52	40	
90	101	55	43	
1700	103	57	45	
10	103	57	45	
20	105	59	47	
30	106	60	48	
40	109	63	51	
50	110	64	52	
60	111	65	53	
70	114	68	56	
80	115	69	57	
90	117	71	59	0
1800	122	76	64	5
10	130	84	72	13
20	136	90	78	19
30	150	104	92	33
40	154	108	96	37
50	161	115	103	44
60	168	122	110	51
70	179	133	121	62
80	187	141	129	70
90	194	148	136	77
1900	205	159	147	88
10	225	179	167	108
20	245	199	187	128
30	254	208	196	137
40	272	226	214	155
50	286	240	228	169

*Source: *Random House Dictionary* (1967).

Since two constants determine a straight line, it follows that two constants completely determine the equation of exponential growth. Complex social or natural processes will usually require more than two constants for their accurate representation, so that exponential growth laws should not be able to represent such processes for long periods except in special circumstances.

If the growth exponent is positive, then $P(t)$ will increase indefinitely as time passes. For populations on the earth, this cannot happen, so it must be the case that a population which follows an exponential growth process must ultimately change its rate of growth. This change cannot be described without introducing new assumptions; it is by no means clear what these assumptions should be for human or book populations, although a number of proposals have been made.

If $P(t)$ does grow exponentially, then it doubles every $\log 2/a^0$ units of time; if time is measured in years (as will be assumed from here on), then $P(t)$ doubles every $(0.69315/a^0)$ years. The annual rate of increase is given by $(e^{a_0}-1)$; if a_0 is small, this is approximately equal to $(a_0 + a_0^2/2)$. For instance, if $a_0 =$ 0.1, then the annual growth rate will be 0.1052. . . , just more than 10 percent. For growth rates, or growth exponents, less than 0.1, the growth exponent is essentially the same as the annual growth rate.

Recognition that most growth processes could not be described by a two-parameter curve such as the exponential led many investigators to attempt generalizations having a greater number of parameters that could be adjusted so as to fit the data. This is a difficult problem. There are mathematical theorems which state that any sufficiently smooth curve—and all of the growth curves that we are considering satisfy this condition—can be represented as closely as desired if enough parameters are used. The practical problem is to provide a representation that uses as many parameters as are necessary, but no more, so that there is some hope that the representation actually corresponds to the actual underlying physical or probabilistic processes in a natural manner. The earliest well-reasoned generalization of the exponential to fit growth data was made by Verhulst (Ref. 14) in the mid-nineteenth century; his discovery of the *logistic curve* was independently repeated by Pearl and Reed (Ref. 11) in 1920. Verhulst's idea was to replace the differential equation (1) by the "simplest" generalization. Upon dividing both sides of eq(1) by $P(t)$, one finds

$$\frac{dP\ /\ dt}{P} = a_o\ .$$

Verhulst assumed that the right-hand side should be replaced by a more general function, say $f(t,P(t))$, which might depend on the time t as well as on the population $P(t)$. If the population growth rate depends only on the population, and not on the time (this is realistic; for instance, had the United States been discovered 300 years earlier, its population growth is likely to have proceeded in the same manner as actually occurred) then f depends only on $P(t)$. Verhulst's next assumption was that this dependence could be expressed by means of a power series, that is, in the form

3) $$f(P(t)) = a_o + a_1 P(t) + a_2 P(t)^2 + \ldots\ldots,$$
$$(a_o > o)$$

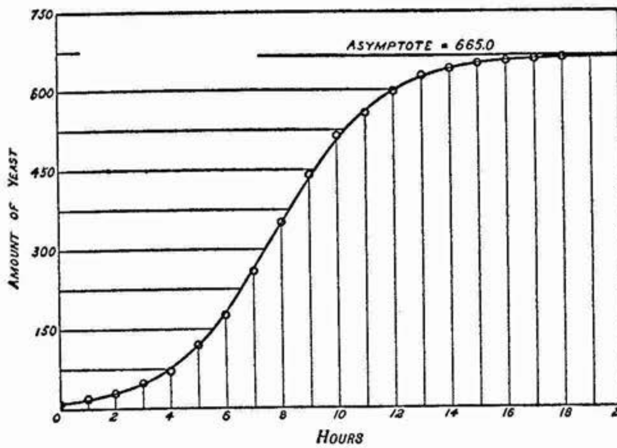
further, he made the approximation that all of the terms on the right except the first two in eq (3) could be neglected, and arrived at the differential equation

4) $$\frac{1}{P}\frac{dP}{dt} = a_o + a_1\ P\ .$$

The solution is the *logistic function*,

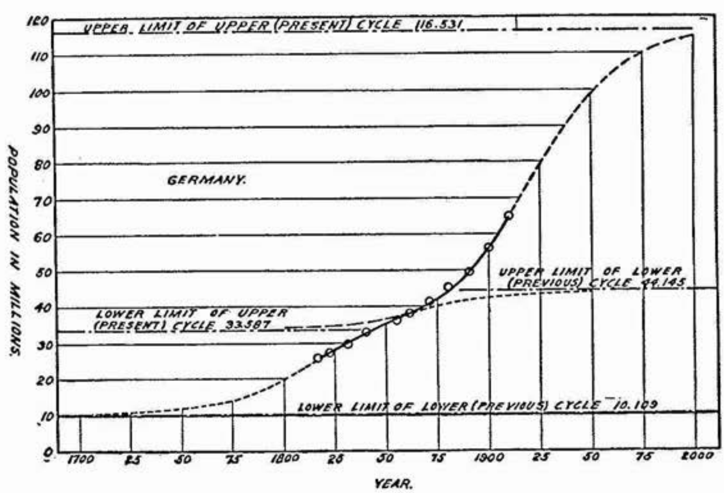
5) $$P(t) = \frac{-a_o / a_1}{1 + e^{-a_o t + c}}\ .$$

where c is a constant which must be determined from the value of $P(t_0)$ at some time t_0. This equation involves three parameters a_0, a_1, and c instead of the two appearing in the exponential. For $P(t)$ to be positive it is necessary that a_1 be negative; then the term $a_1P(t)$ in the differential eq (4) corresponds to an influence retarding the growth of P with time, and in fact, as t becomes indefinitely large, $P(t)$ does not, but rather approaches the value $-a^0/a^1$ as an absolute maximum. The positive number a_0 represents, as before, the growth exponent. Pearl (Ref. 10) provides a useful example of a logistic curve, making use of Carlson's data on the growth of yeast cells; we have reproduced this curve in Figure 6. It illustrates both the shape of the logistic and the fact that logistic curves do sometimes provide accurate representations of growth processes. Unfortunately, the complex processes of library imprint growth, human population growth, or other



$$P(t) = \frac{66.5}{1 + e^{-0.5355t + 4.1896}}$$

FIGURE 6. The Logistic Curve: Growth of a population of yeast cells, taken from R. Pearl (Ref. 10).



For the period up to 1855, $r(t) = 10.109 + \frac{34.036}{1 + 2.495e^{-0.0394t}}$

From 1855 on, $P(t) = 33.587 + \frac{82.944}{1 + 297.546e^{-0.0472t}}$

FIGURE 7. The population growth of Germany, showing two cycles of growth which have overlapped during the period of census history.

growth phenomena related to society do not in general behave logistically. For instance, neither the holdings growth of the Library of Congress (Figure 1) nor the growth in the number of universities (Figure 5) are well represented by logistic curves. Numerous attempts have been made to fit the data of civilization to logistic curves and to successions of logistics (cf. Ref. 2, 6, 7, 9, 10, 11, and 16).

If it is found that some data can be nicely fit by a sequence of N logistics, this means that at least $4N - 1$ parameters are involved (as well as some additional ones to describe where consecutive logistics are to be fit together, but this can be ignored since this problem is common to all piecewise fitting processes), since each logistic requires three parameters, giving $3N$, and all but one (and sometimes that last one also!) must be shifted up or down, which requires an additional constant. For instance, if data can be fit by two logistics, as shown in Figure 7 taken from Pearl (Ref. 10), then eight parameters are required. Only 11 data points were available to Pearl, so it is no surprise that an eight-parameter function could be found that would provide a good fit; it is not clear that other functions might not provide an equally good fit with the use of fewer parameters. Indeed, Pearl's data is shown on a semilogarithmic scale in Figure 8, from which it is readily seen that the leftmost four points are well fit by the two-parameter exponential, the next three could be fairly well fit, and the remaining four are again well fit by an exponential. Therefore three exponentials, requiring a total of six parameters, appear to do about as well as two logistics requiring eight parameters. One difference is that there is no evident place to transfer from one logistic to the other; Pearl arbitrarily does this at 1855. The exponential fits of Figure 8 immediately suggest that something happened between the adjacent data points for 1840 and 1855, and the fact that the lines representing the earliest and latest exponentials in the graph are nearly parallel suggests that whatever occurred to change the population growth rate between 1840 and 1855 had returned to "normal" by 1870. The revolutions and turmoil of 1848 and the following years could have affected the birth rate, and it might have taken a generation—about 20 years—to recover the rate loss, thus accounting for all of the features of this graph in an informative way that the logistic interpretation does not permit. Indeed, it is precisely the *fluctuation* from exponential growth that is of interest in this case; the logistic curves smooth that fluctuation so as to make it invisible.

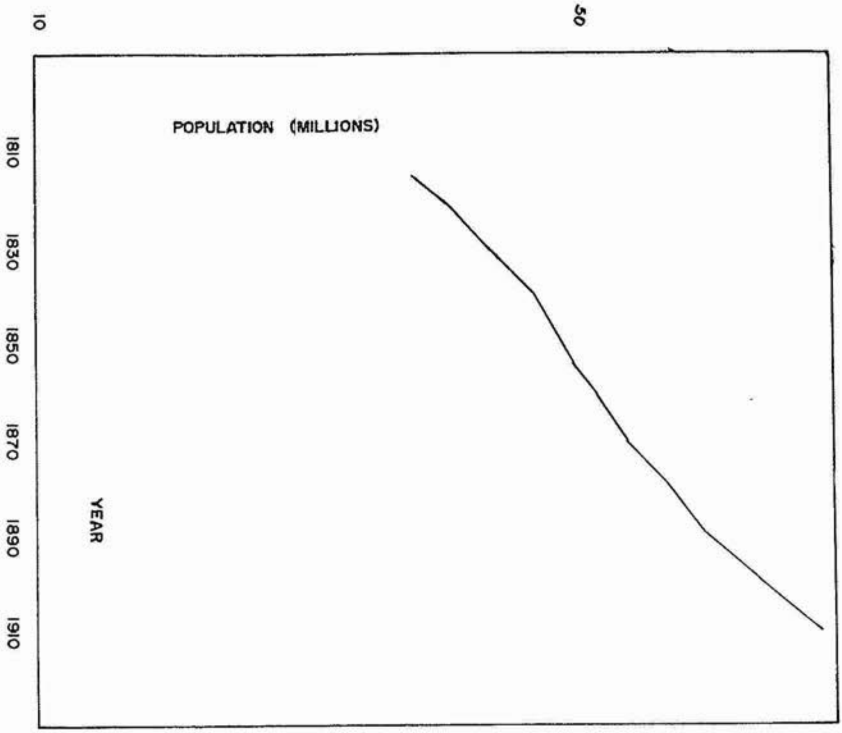


FIGURE 8. The population growth of Germany 1816-1910.

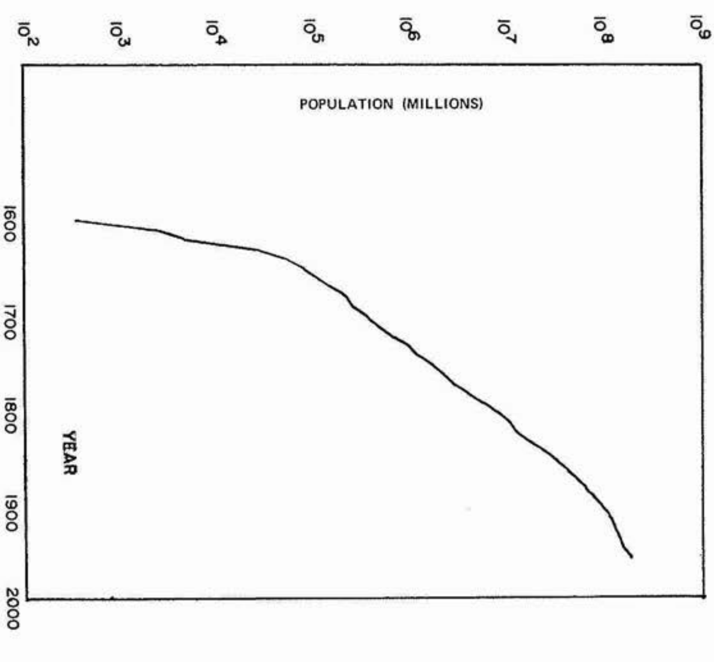


FIGURE 9. United States Population 1610-1965.

The following paragraphs will argue that *piecewise exponential* approximation to growth curves provides the most convenient and informative tool for understanding the underlying processes. We will argue that the nature of civilization is such that when influences which tend to retard growth are encountered, the *exponent* of the exponential growth curve is changed to relieve that retarding pressure; in general, the *form* of the function is not changed. This is the same as saying that the growth exponent of the exponential is a function of time, and, with a sufficiently complex variation of this exponent with time, any growth curve can be traced out. But it appears from a study of various important classes of growth curves that the variation of the exponent with time is of a particularly simple nature. It is constant most of the time but, in reaction to external retarding influences, it makes a transition from its original constant value to another constant value which is compatible with the external circumstances.

Consider, for instance, the population growth of the United States, shown in Figure 9, for the period 1610 to date. Until 1690, growth was extremely rapid and exponential; from 1700 until 1880 growth was again exponential with a remarkable degree of accuracy and consistency. After 1880 a more complex pattern occurs (to be discussed in Section 4). The departure from exponential growth which shows up for the first time in the 1880 census corresponds to what historians call "Turner's thesis," concerned with the closing of the Western Frontier. It is evident that the 270-year period from 1610 until 1880 can be accurately described by two exponentials, thus four parameters.

Population statistics are unusually complete for the United States; most other nations have only recently begun to accumulate accurate population estimates based on census data. Nevertheless, it will be useful to look at the population growth of one other nation. Figure 10 displays the growth of Japanese population since 1872. It can be accurately represented by two exponentials, from 1872 until 1900, and from 1900 to date. There is a sharp population decline, shown by the 1945 census, due to manpower losses in World War II, but this was completely made up by the time of the 1950 census. The long-term trend rate of Japanese population growth has returned approximately to that level it has held since 1900. In the popular press as well as in demographic literature there has been concern for the rapid growth of the Japanese population since the war, and some relief that it appears to be coming back under control at last. This is misleading; in fact, having retrieved

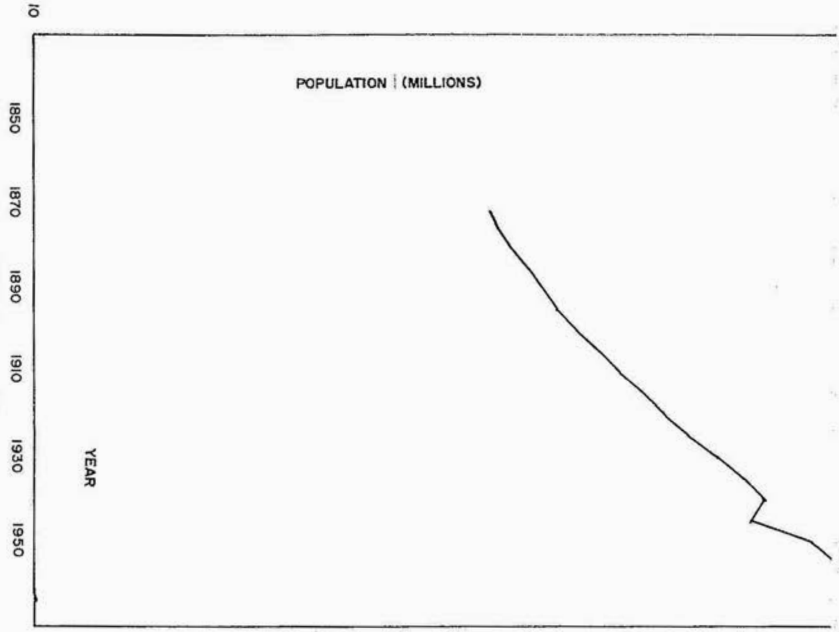


FIGURE 10. Jananese population.

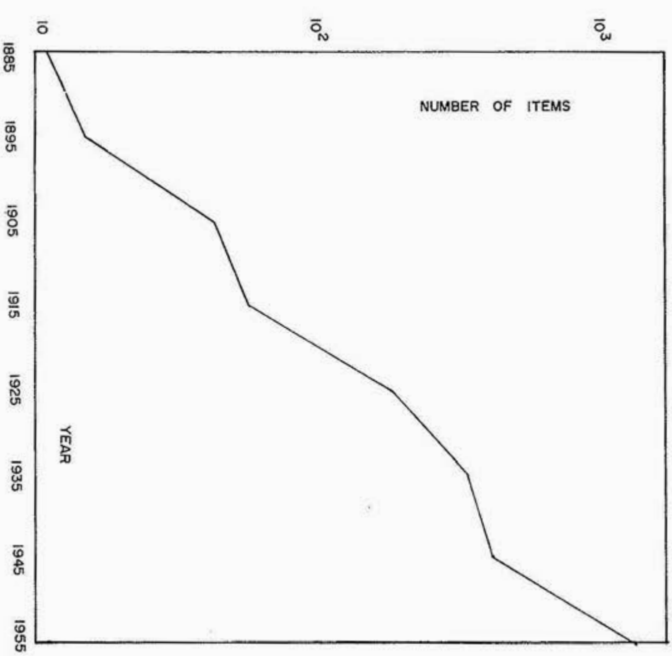


FIGURE 11. Imprint date distribution from the Stanford Undergraduate Catalog 1966.

the population lost in the war, the growth rate is now returning to its traditional value. This shows a danger inherent in looking at short sections of time series. The local fluctuations may obscure the general picture and lead to gross misestimates of what is happening. Further examples of this type of problem will come up in what follows.

Returning to Figure 2, the Widener Shelf List Volume 7 imprint date distribution, observe its essential exponentiality. Figure 11 displays the same information for the Stanford Undergraduate Library. This collection is restricted to recent imprints because of its small size and limited purpose. Nevertheless the usual exponential trend is clearly evident. It appears to make no difference whether a subcollection of a large library or an entire small library is examined. In Reference 3 it was shown that similar growth occurred in a random sample from a university library of some 300,000 items, exclusive of periodicals.

Figure 12, taken from Reference 2, shows that the number of scientific periodicals and also the number of scientific abstract journals have been growing exponentially, the former for approximately 300 years.

Turning to quite a different type of growth statistic, in recent years the Basic Oxygen Process has been increasingly used for the production of steel in the United States. Growth in the output of BOP raw steel is shown in Figure 13. It consists of two parallel lines on the semilogarithmic graph paper, separated by about a year. Thus, apart from this fluctuation (which will be discussed in Section 4), this statistic also follows the exponential growth law.[3]

4. *Local Fluctuations in Growth*

All of the graphs that have been discussed show consistent exponential trends for most of their duration, but there are deviations of several types.

I. There are minor fluctuations which appear to have an average value of zero with respect to the underlying exponential trend. These probably correspond to random influences that are of no long-range importance, and which cannot be subjected to a deterministic analysis. There is not much that has to be said about them other than that they always exist in natural time series and that there is little that can be done to analyze them. Figure 14 illustrates the residuals of the Widener Shelf List data used for Figure 2 with respect to exponential trend lines. Trends were obtained using least

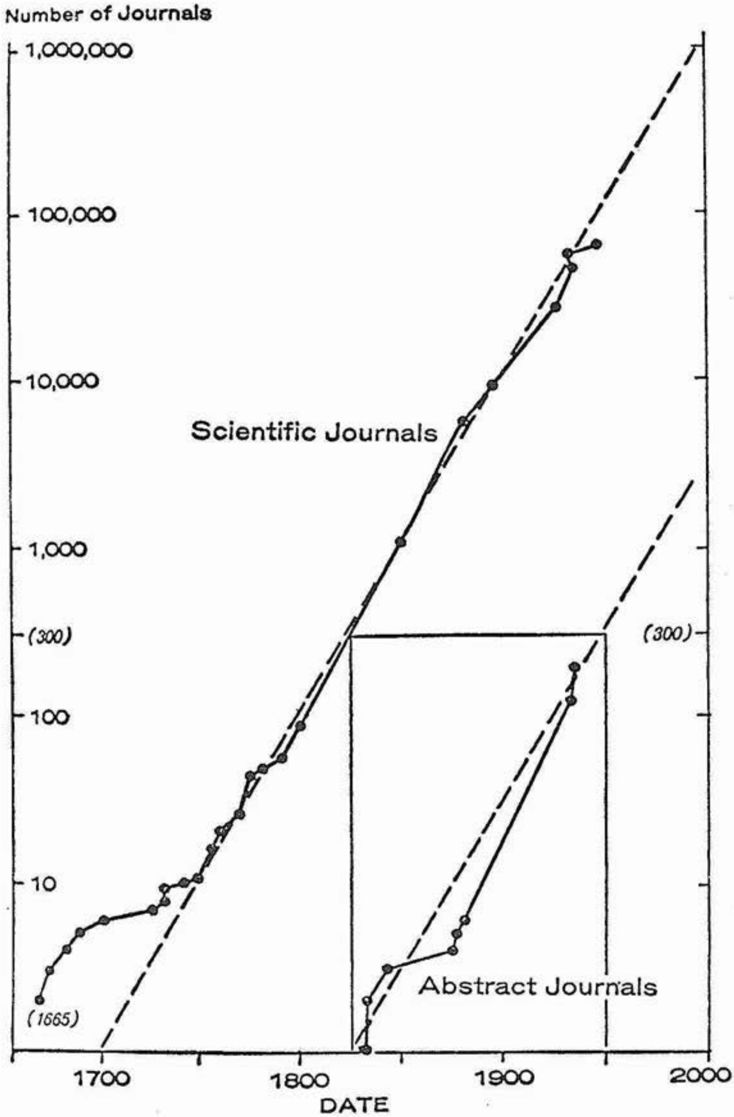


FIGURE 12. Number of Scientific Periodicals (Data from D. J. de Solla Price, *Science since Babylon* [New Haven, 1961], p. 97).

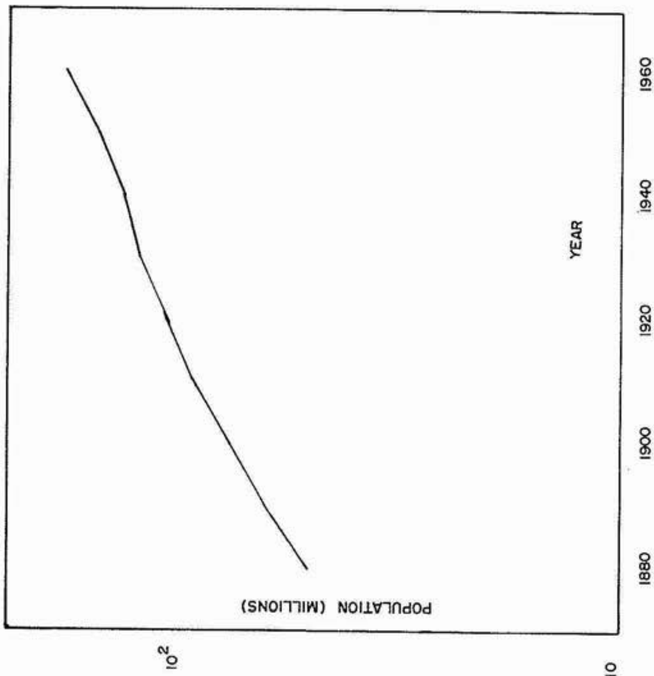


FIGURE 15. U. S. Population 1880-1960.

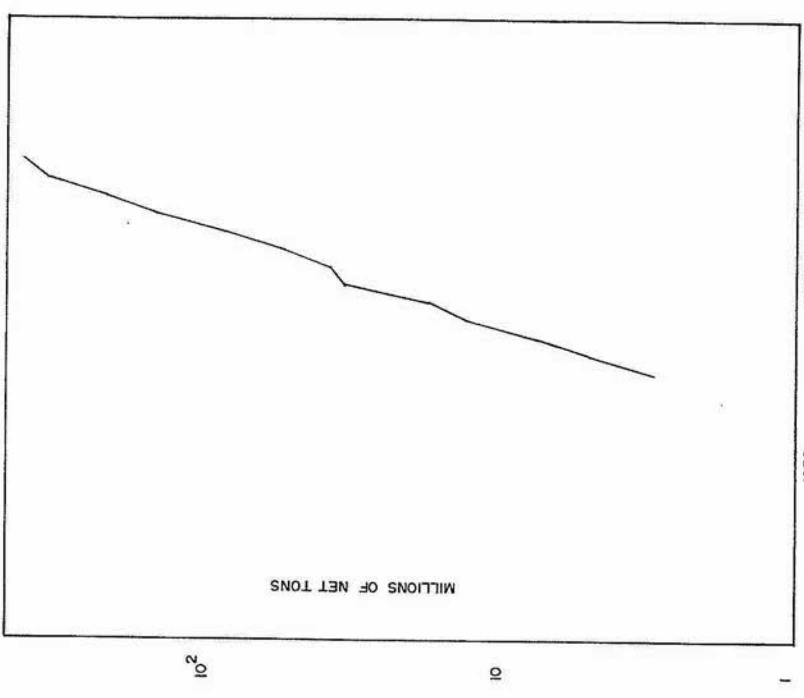


FIGURE 13. Basic Oxygen Process – Raw Steel Output in the United States.

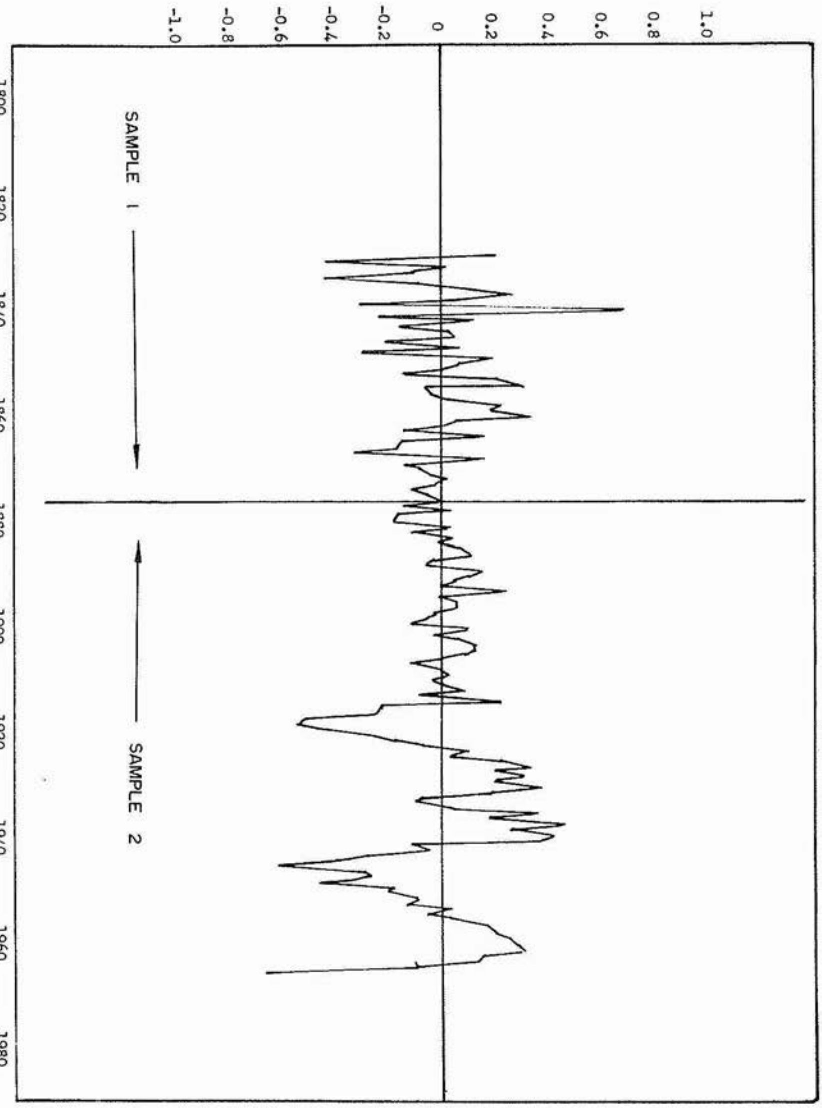


FIGURE 14. Widener Residuals Volume 7: Imprint distribution least squares residuals.

squares methods on the logarithms of the data for two distinct samples: the period 1830-1876, and the period 1876-1965. The residuals are the deviations of the logarithms of the data from the logarithms of the fitted exponentials. From 1870 to 1914 it appears that the residuals are essentially random; this represents Type-I fluctuation. The three largest residuals occur in 1965, 1945, and 1918. The first is due to the incompleteness of the collection in the most recent years; plotting by imprint date invariably introduces a bias in the most recent figures because items published in any given year are acquired over a span of years following. Technically, this bias extends over the entire collection; however, it is most noticeable in the most recent five to ten years. It would be of interest to determine the distribution of imprint dates for a given year's acquisitions to determine the effect of the bias more precisely. The other two large residuals occur in the final years of the world wars, and are both negative, as might be anticipated. These fluctuations are clearly not Type-I. The large residuals that occur in early portions of the sample (e.g., that in 1840) are of questionable significance because of the small sample sizes for those early years.

II. There are departures from exponential growth which last for a short period relative to the duration of the exponential part, followed by a longer period of stable exponential growth proceeding at a different growth rate. This type of departure from exponential regularity appears to correspond to a change in the underlying environment which requires a readjustment of the growth rate. The readjustment is effected by passing in some (possibly irregular) manner from the initial constant growth rate to the new constant rate by means of a transition period of relatively short duration. Two examples of this type of process are shown in Figure 9 (U. S. Population), one in Figure 10 (Japanese Population), and two in Figure 1 (LC holdings).

III. The third type of fluctuation is perhaps the most interesting. It is represented by a period of regular exponential growth which is followed by a transition period of relatively short duration. The transition period precedes another period of stable exponential growth (as in Type-II above) which proceeds at the *same rate of growth as the exponential preceding the transition.* Thus, on semilogarithmic graph paper,

> the initial and final exponentials will appear as *parallel* lines; the transition corresponds to a curve connecting the two exponentials which can be irregular. The two exponentials may actually be parts of one; this is shown in Figure 10, where the transition is constituted by the rapid decrease in Japanese population from 1940 to 1945 followed by an even more rapid increase from 1945 to 1950, bringing the population back up to the exponential trend line as if World War II had not occurred. In other cases, the two parallel lines representing exponential growth do not coincide; usually, the terminal line lies below the initial one, which means that although the population growth rate has returned to its initial state, there has been an unrecovered absolute loss in population. This is well illustrated by Figure 13 (Basic Oxygen Process Raw Steel Output in the U. S.) which shows an unrecovered loss which can be interpreted as having set the industry back by slightly more than one year.[4]

There are other types of departure from exponential growth, but they are not so easily characterized nor do they seem to play an important role in the types of populations that are under consideration in this paper.

Fluctuations of Type-III and their connection with unrecoverable losses are worth some further discussion. An important example is furnished by population and economic time series for the United States encompassing the period of the Great Depression, which is a Type-III fluctuation.

Figure 15 shows the population of the United States during the 1880-1960 period, as given by the decennial census. In connection with Figure 9, it was pointed out that a Type-II transition occurred about 1880, corresponding to the closure of the frontiers; the effect of this transition is visible in the first three data points of Figure 15; the next three lie on a line (which is not remarkable since two points determine a line). The remaining three points lie on another line which is nearly parallel to the first. Therefore, the data from 1910 to 1960 can be interpreted as indicating a constant growth rate with a Type-III fluctuation occurring between 1930 and 1940, the decade of the depression. Pursuing the implications of this interpretation, shift the 1940-1960 line to the left so that it coincides with the 1910-1930 line, and measure the number of years of shift required to obtain this coincidence: it is approximately 6 years (we will use 5.89 years). Figure 16 shows the same data used for Figure 15 with a 5.89-year shift to the

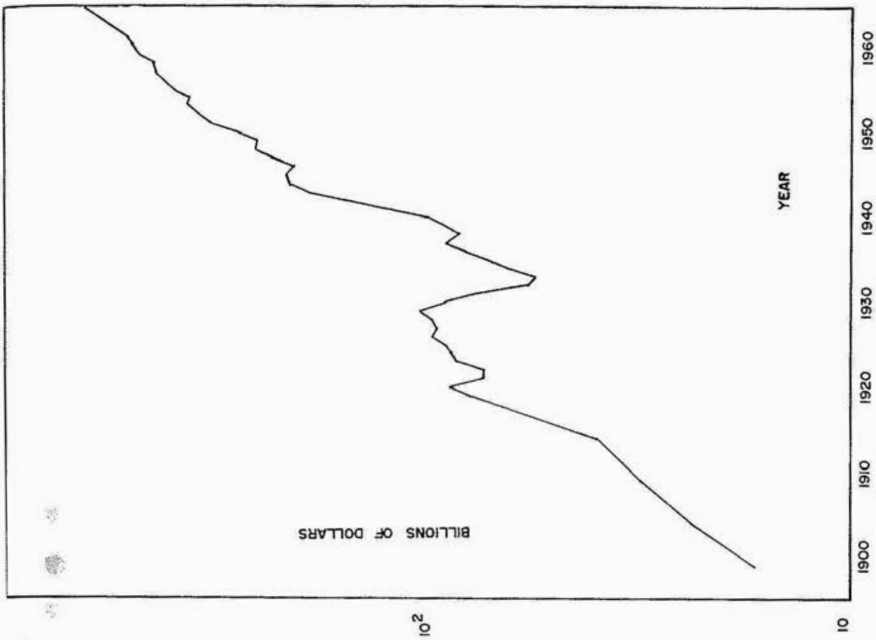


FIGURE 17. U. S. Gross National Product in current dollars.

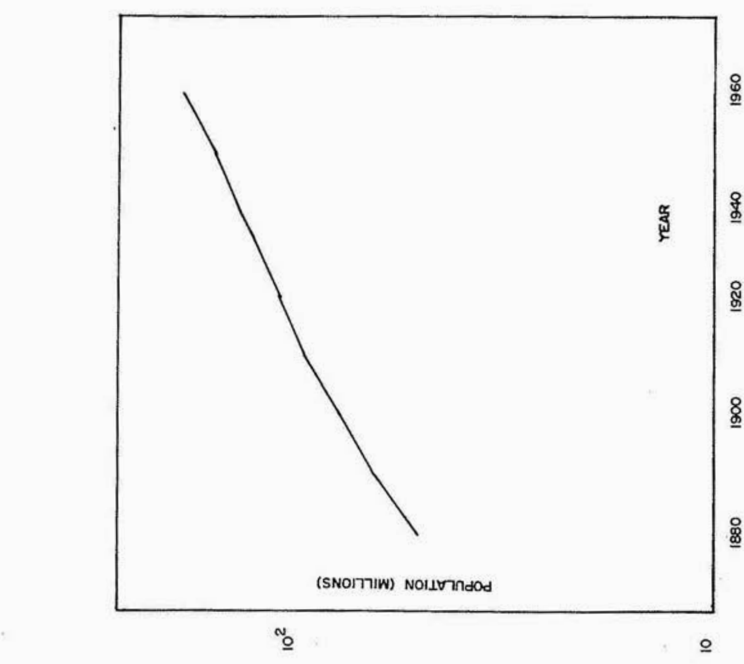


FIGURE 16. U. S. Population 1880–1960 with a 5.89 year correction.

past for post-1930 data. The six points corresponding to the population figures from 1910 through 1960 now lie on a line with very good precision. It is possible to interpret this Type-III fluctuation as implying that the *effective duration of the Great Depression was about 5.89 years*. In other words, the non-recoverable population loss due to the depression was equivalent to 5.89 years of exponential population increase at the previously and subsequently prevailing rate. It should be observed that World War II had as little effect on the U. S. population growth rate as it had on that of the Japanese population (cf. Figure 10).

As a check on the notion that there was really a non-recovered loss of population during the Great Depression, we can investigate the behavior of the Gross National Product (GNP) on the assumption that there is a close relation between the two time series. This can be done either by assuming the 5.89-year gap and testing the departure from linearity using this gap, or by deriving the gap (if any) from the GNP data directly.

Figure 17 shows the GNP (in current dollars) for the period 1894 to 1965. The values up to 1921 are 5-year averages; the subsequent values are by individual years. The data prior to World War I and the data subsequent to World War II are quite consistent with the general hypothesis of exponential growth. The intervening period (1914-1945) exhibits rather wild fluctuations—as might be expected. From this superficial examination alone it is quite clear that GNP does not possess the stable growth pattern shown by population growth.

Figure 18 shows the same data with the 5.89-year interval in the 1930's removed. Even with the larger variation of the GNP data, it is clear that this data is not inconsistent with the "depression-gap" hypothesis. In fact, a good portion of the data for the 1920's is included within the rough limits of variation sketched in Figure 18. One could conjecture that the real roots of the Great Depression are to be found in the "excessive" growth of the GNP in World War I, and in the difficulty in guiding the GNP back to its basic growth rate without overshooting the goal.

Derivation of the depression-gap value from GNP alone leads to some problems. If one uses the 5-year averages for the two periods of evident linearity before World War I and after World War II, one obtains a depression gap of 2.05 years. This can be explained in part by the variation introduced by the Korean War; if the 5-year period including the Korean War is eliminated, the estimate of the gap becomes 3.29 years. Better agreement might be

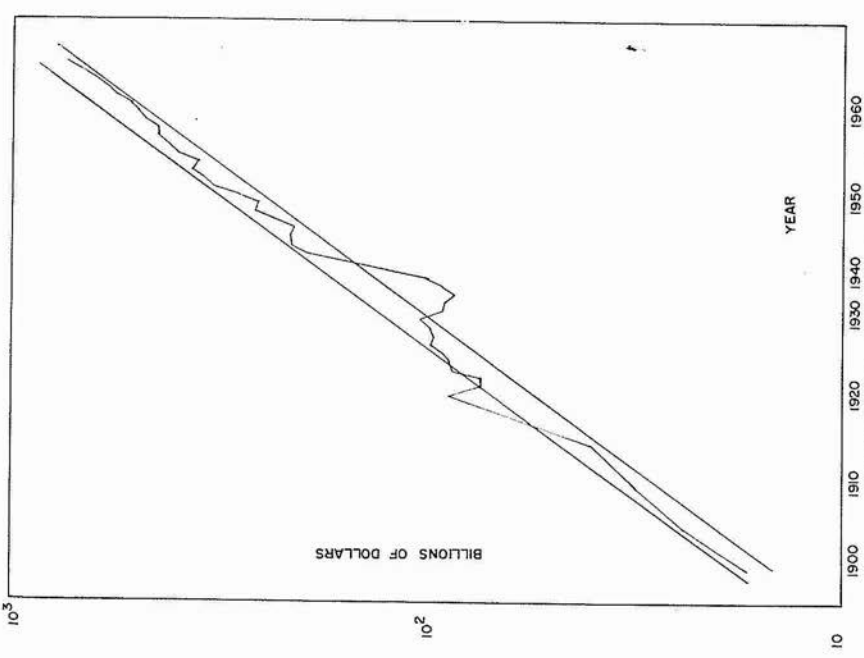


FIGURE 18. U. S. Gross National Product in current dollars with a 5.89 year correction for the Depression.

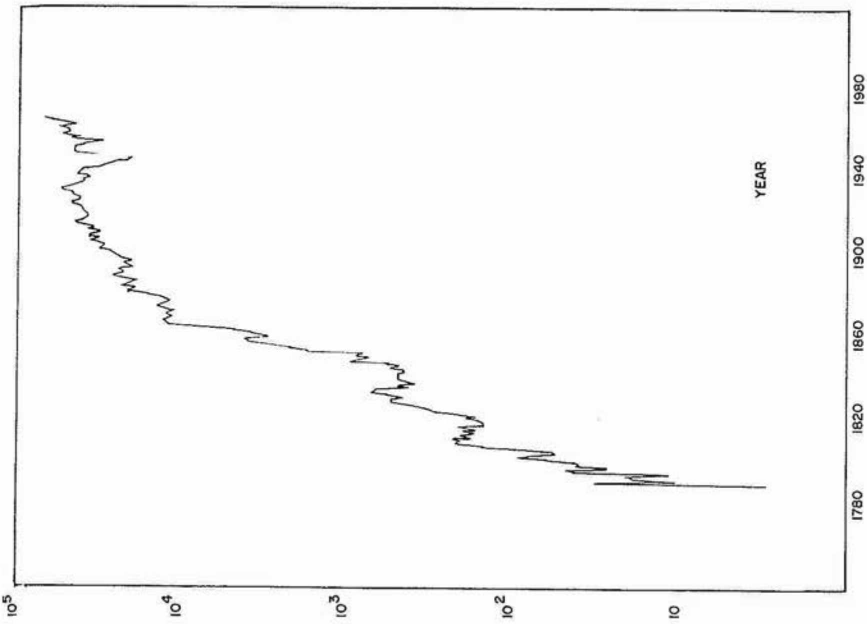


FIGURE 19. Inventions, U. S. Patents issued 1790–1966.

obtained were it possible to obtain GNP figures for individual years prior to World War I or if the study were based on real rather than on current data.

Our main interest in this data, however, is in showing the need for careful analysis of long-term growth as well as growth in the short term. Without sufficient long-term statistics, a Type-III fluctuation may be misinterpreted as one of Type-II, and an underestimate of the trend growth rate may be made. Indeed, this has often happened in the past 30 years; libraries and other institutions have rapidly outgrown facilities constructed under the misapprehension of future lower trend growth rates due to an analysis of insufficient portions of their growth records.

Figure 19 exhibits the history of United States invention patents issued from 1790 through 1966 (cf. Table III). This complex graph contains fluctuations of all three types. There is an important Type-III fluctuation during the Great Depression which increases in its effect during World War II, and there are two others, from about 1812 to 1822, and 1830 to 1845. There is a basic change in the trend growth rate, a Type-II deviation, at about 1870 which appears to correspond to the Type-II transition observed in the U.S. population curve (Figure 9) about 1880. And there are the inevitable Type-I random fluctuations throughout the graph.

5. *Conclusions*

It appears to us that future research of a more complete and detailed nature will bear out the following conclusions:

1. Exponential growth of library holdings will persist for the foreseeable future. To maintain current growth rates, automation of the production of portions of the intellectual content, as well as the production of books and equivalent forms of stored information, will increase.

2. Growth curves of importance to library management consist in general of piecewise exponential segments connected by transitional fluctuations. Determination of the nature of the current and short-term future portions of the curves is necessary for realistic and practical planning purposes. Piecewise exponential approximations are the simplest techniques for exhibiting the structure of these growth curves.

3. As far as growth *rates* are important for planning purposes, fluctuations of Type-III must be detected and *ignored* in the evaluation of future requirements of a library.

TABLE III

U.S. INVENTION PATENTS ISSUED

1790 - 1966

Date	Number	Date	Number	Date	Number	Date	Number	Date	Number
1790	3	1830	544	1870	12137	1910	35141	1950	43040
1791	33	1831	573	1871	11659	1911	32856	1951	44326
1792	11	1832	474	1872	12180	1912	36198	1952	43616
1793	20	1833	586	1873	11616	1913	33917	1953	40468
1794	22	1834	630	1874	12230	1914	39892	1954	33809
1795	12	1835	752	1875	13291	1915	43118	1955	30432
1796	44	1836	702	1876	14169	1916	43892	1956	46817
1797	51	1837	426	1877	12920	1917	40935	1957	42744
1798	28	1838	514	1878	12345	1918	38452	1958	48330
1799	44	1839	404	1879	12125	1919	36797	1959	52408
1800	41	1840	458	1880	12903	1920	37060	1960	47170
1801	44	1841	490	1881	15500	1921	37798	1961	48368
1802	65	1842	488	1882	18091	1922	38369	1962	55691
1803	97	1843	493	1883	21162	1923	38616	1963	45679
1804	84	1844	478	1884	19118	1924	42574	1964	47376
1805	57	1845	473	1885	23285	1925	46432	1965	62857
1806	63	1846	566	1886	21767	1926	44733	1966	68406
1807	99	1847	495	1887	20403	1927	41717		
1808	158	1848	583	1888	19551	1928	42357		
1809	203	1849	984	1889	23324	1929	45267		
1810	223	1850	883	1890	25313	1930	45226		
1811	215	1851	752	1891	22312	1931	51756		
1812	238	1852	885	1892	22647	1932	53458		
1813	181	1853	844	1893	22750	1933	48774		
1814	210	1854	1755	1894	19855	1934	44420		
1815	173	1855	1881	1895	20856	1935	40618		
1816	206	1856	2302	1896	21822	1936	39782		
1817	174	1857	2674	1897	22067	1937	37683		
1818	222	1858	3455	1898	20377	1938	38061		
1819	156	1859	4160	1899	23278	1939	43073		
1820	155	1860	4357	1900	24644	1940	42238		
1821	168	1861	3020	1901	25546	1941	41109		
1822	200	1862	3214	1902	27119	1942	38449		
1823	173	1863	3773	1903	31029	1943	31054		
1824	228	1864	4630	1904	30258	1944	28053		
1825	304	1865	6088	1905	29775	1945	25695		
1826	323	1866	8863	1906	31170	1946	21803		
1827	331	1867	12277	1907	35859	1947	20139		
1828	368	1868	12526	1908	32735	1948	23963		
1829	447	1869	12931	1909	36561	1949	35131		

4. Much more detailed studies of the time series associated with library operations must be made. Analytical methods that will permit the objective determination of the points of transition from periods of one type of growth to another must be developed as well as the interrelation of statistical information pertaining to different types of growth statistics.

5. It seems clear that there is a connection between the growth of the library archives and the growth of various estimators of the state of civilization. It ought to be determined which is cause and which is effect. More precisely, the role of the preservation and transmission of information in the development of civilization certainly ought to be investigated not only as a subject of abstract intellectual interest but also to provide a working tool for those responsible for allocation of national resources. Furthermore, as the details of this relationship become clarified, library management will be in a better position to improve the means of accessing the information archive.

NOTES

1. Gelb (Ref. 4) writes: "If the alphabet is defined as a system of signs expressing single sounds of speech, then the first alphabet which can justifiably be so called is the Greek alphabet."

2. The data given in Table II appears to agree with that used by De Solla Price for his figure on page 115 of Reference 2, but his description of the growth there given seems to be in error.

3. The datum of 1967 does not lie on the fitting line. Since the most recent statistic in an economic time series is usually revised, this number has been ignored in fitting the exponential.

4. The Kennedy confrontation with the steel industry concerning its pricing policies occurred *after* the sharp transitional drop in BOP steel output growth rate in 1960-61. The proposed price increases may have been a reaction to this transition. Had it been known at the time that the transition was of Type-III but not of Type-II and that it would last for one year in its depressive phase, the industry might not have reacted with its proposed non-transitional price increase. This example illustrates the importance of studying the causes of transitions and learning how to distinguish the various types as they occur.

This problem also illustrates the importance of further statistical study of the procedures to be used in interpreting growth data. It is possible to "fit" the BOP steel output data with a single straight line that effectively "hides" the transition period of 1960-61. The procedures for testing the improvement introduced by the notion of a break in the growth function are fairly obvious. Isolation of the period involved, given that a break occurred, is also straightforward. However, the problem of finding break points when there is no a priori information as to how many are present does not appear to have been studied at any length.

REFERENCES

[1] AMERICAN IRON AND STEEL INSTITUTE, *Charting Steel's Progress During 1967* (New York, 1968).

[2] DE SOLLA PRICE, D. J., *Science Since Babylon* (New Haven, 1961).

[3] DOLBY, J. L., V. J. FORSYTH, and H. L. RESNIKOFF, *Computerized Library Catalogs: Their Growth, Cost, and Utility* (Cambridge, Mass., 1969).

[4] GELB, I. J., *A Study of Writing* (Chicago, 1952).

[5] GRANGER, C. W. J., and M. HATANAKA, *Spectral Analysis of Economic Time Series* (Princeton, 1964).

[6] HART, HORNELL, "Logistic Social Trends," *American Journal Sociology*, 50 (1945), 337-352.

[7] ———, "Depression, War, and Logistic Trends," *American Journal Sociology*, 52 (1946), 112-122.

[8] HENDERSON, J. W. and J. A. ROSENTHAL, *Library Catalogs: Their Preservation and Maintenance by Photographic and Automated Techniques* (Cambridge, Mass., 1968).

[9] PEARL, R., *Studies in Human Biology* (Baltimore, 1924).

[10] ———, *The Biology of Population Growth* (New York 1930).

[11] PEARL, R., and L. J. REED, "On the Rate of Growth of the Population of the United States Since 1790 and its Mathematical Representation," *Proceedings National Academy of Science (USA)*, 6 (1920), 275-288.

[12] *The Statistical History of the United States from Colonial Times to the Present* (Fairfield Publishers, Inc., 1965).

[13] STRUIK, DIRK J., *A Concise History of Mathematics*, 3rd edition revised (New York, 1967).

[14] VERHULST, P. F., "Recherches Mathématiques sur la Loi d'Accroissement de la Population," *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18 (1845), 1-38.

[15] *Widener Library Shelf List (Volume 7): Bibliography and Bibliography Periodicals*, Harvard University Library (Cambridge, 1966).

[16] YULE, G. UDNEY, "The Growth of Population and the Factors Which Control It," *Journal of the Royal Statistical Society*, 88 (1925), 1-58.