sion upon the radius would leave 10% with more than 4 times the average radius and 10% with less than one fourth of that value. Assuming the existence of Eddington's high density stars, the last alternative of large dispersion in radius presents no theoretical difficulties, the only practical objection is that not one of such small, high-density O stars has been found.

Allowing for the limitations of a discussion like the present, it appears reasonably certain that among the O stars occurs a fairly large dispersion in intrinsic brightness and a fairly large scattering through the depths of space. A decision, however, cannot be reached until we have reliable proper motions for the very faint O stars, or until we have, under the additional assumption that the same physical conditions prevail in the Magellanic Clouds as in the Galactic System, definite information about the absence or presence in the Magellanic Clouds of O stars of the twentieth magnitude and fainter.

THE POPULATION OF NEW YORK CITY AND ITS ENVIRONS

BY EDWIN B. WILSON AND WILLEM J. LUYTEN

HARVARD SCHOOL OF PUBLIC HEALTH AND HARVARD COLLEGE OBSERVATORY

Read before the Academy November 11, 1924

Four years and a half ago Mr. Raymond Pearl read before this Academy the first of the now famous papers by him and his colleague, Mr. L. J. Reed, on following the past course of a population and forecasting its future growth by fitting the curve $1/P = A + Be^{-mt}$ of autocatalytic (or buffer) reaction to the past census counts and extrapolating to predict the probable future course of the population's growth.¹ The two most ambitious subsequent papers are perhaps that presented at the general meeting of the American Philosophical Society last spring discussing the population of a great many countries,² albeit with the aid of a generalized type of curve, and the elaborate discussion of the future population of Greater New York City prepared for the Committee of Plan of New York and Its Environs and dealing with many questions of detail significant for long range city planning.³ The problem of estimating the future population of the Metropolitan district of Boston in connection with planning for an adequate water supply was brought to us at the Harvard School of Public Health some time ago, and at once recourse was had to the Pearl-Reed method as the only one which makes real pretention to scientific forecasting as against merely personal guess-work. On investigating this problem a number of points seemed to thrust themselves forward for consideration in connection with the general method and they may for the present purpose be better

illustrated by commenting on the results already published for New York City than by giving figures for Boston.

The total area which is included for regional planning by the Committee on Plan of New York and Its Environs and for which population predictions are made embraces over 5500 square miles; it is part of the three states, New York, New Jersey and Connecticut, and is, so to speak, New York City proper and its chief commutation areas. This region had in 1920 a population of 9 millions which is 60% of the population (15 millions) of the three states and $8\frac{1}{2}\%$ of the total population of the United States. The predicted ultimate (asymptotic) population of the region is 35 millions or $17\frac{1}{2}\%$ of that predicted by the same method by these authors for the United States¹ (197 millions). What may be the political, economic, or social consequences of such a concentration of population in and immediately

around New York City, or the influence on the public health, and what the system of transportation and distribution whether of persons or of the necessary goods, are questions to intrigue and tax the imagination. Certainly if there is to be such crowding it is well that present and future plans for the city shall take it into consideration long in advance, and should further consider all accessory information available.⁴

Among the items of interest is the condition immediately surrounding the greater city, say, in the three states in which it lies. What is to be their population? Figures for the populations of New York, New Jersey and Connecticut at each decennium from 1790 to 1920 are available in the Census Reports. A curve of the type $1/P = A + Be^{-nt}$ may be fitted to the data, namely, $1/P = 0.045 + 1.264 e^{-0.0302t}$, where t is

Population of Greater New York 1790-1920 (lower curve) with two extensions after 1920 corresponding to the Pearl-Reed and the Wilson-Luyten fittings of the curve, asymptotic respectively at 35 and 17 millions. Population of New York, New Jersey and Connecticut 1790-1920 and extended, asymptotic at 22 millions (upper curve).

measured from 1790 in years. Judged by its percentage departure from the recorded census figures the fit is not bad.⁵ In figure 1 are given the census

data for the states and the fitted curve carried on into the future, and the same for the region. The Pearl-Reed limiting population for the region is about 35 millions whereas our limiting value for the three states containing it is a little over 22 millions. The curves cross just before the year 1960. If we follow the curves for 40 years after the recent census, we have the population of the part exceeding that of the whole. As the three states outside the confines of New York City and its environs contained 6 millions in 1920 and the curve of this exterior population shows no signs of falling off (but apparently still tends to increase), the essential conflict between the two curves of growth is emphasized.

This conflict is not surprising. The authors specifically state that the curve can be trusted to forecast only so long as growth remains in the same cycle and no new factors or forces intervene;⁶ in their treatment of Germany they break the growth into two cycles. Insofar, therefore, as New York City and its environs on the one hand and the group of the three states of New York, New Jersey and Connecticut on the other hand may be subject to treatment by the simple Pearl-Reed law of population growth and insofar as the curves which they and we, respectively, have fitted may be trusted adequately to represent the growth, we may say that one or the other or both of these populations must very soon change their cycle of growth, if in fact they have not already done so.⁷

The conflict, possibly only to a limited extent so far as practical results go, is logically inherent in the method employed. Populations in different regions are necessarily additive to form the population of the whole. The older types of curve fitted to population growth were additive⁸ but had the serious biological defect of giving infinite populations at infinite time and thus failing to predict even qualitatively the inevitable saturation of a region with population. The Pearl-Reed curves, both the simplest one here used and the more complicated ones, satisfy (sometimes⁹) the biological condition for ultimate saturation but fail to satisfy the logical condition of additivity and hence when applied to different parts of a population may well make the limiting population of a part greater than that of the whole.¹⁰ One may as well disregard a logical as a biologic requirement; but in either case the validity of the law established is subject to question, and the use of the law for the prediction is more than ordinarily hazardous.

Although the logical conflict would remain, it seemed worth while, on account of the apparently systematic departures of the populations from the fitted curves, to try a fit by the generalized skew growth curve suggested by the authors and applied by them to Germany and Japan.¹¹ The first step in making the fit is to determine the limiting population by equation xxvii of p. 577 of *Studies in Human Biology*. Using the enumerated census of New York City and its environs as given by Pearl and Reed³ at

the five equidistant censuses of 1790, 1820, 1850, 1880 and 1910, an equation was obtained which had no root in the neighborhood of 35 millions, their predicted asymptotic population, and it thus appeared that their general curve could not in this case even approximately reproduce the future results given by the simple curve.¹²

It was a succession of disappointments of this sort, an unexpected instability of the forecasts as obtained by different systems of weighting and by different types of the Pearl-Reed curves that led us to place little confidence in our own ability to apply their method to predict satisfactorily the future population of Greater Boston which must depend for its water supply upon the foresight of Metropolitan District Commission. It was for the purpose of getting light on the method that the examination of the figures for New York was made, and the results of that examination, partly detailed above, leave us skeptical of the forecasts for New York. It is to be hoped that further developments, applications and explanations by the authors will serve to allay our doubts which at present seem shared by a great many economists.

One further difficulty of a mathematical nature may be mentioned. As the curves are not additive the sum of the squares of the residuals, which is the quantity to be minimized in a least squares solution, is not quadratic in the variables (the parameters of the curve), and the derivatives are not simple linear, but complicated transcendental, equations in those variables. It is therefore practically hopeless to work analytically and becomes necessary to explore the whole ground by tedious arithmetic trials without any assurance that there may not be several minima instead of a single value. Indeed when some of the errors are so large as they must be in any non-sinusoidal curve fitted to the New York population one might expect that the solution would not be unique.

This may be illustrated as follows: The general curve with which we are working is $1/(P - d) = A + Be^{-nt}$ containing four constants. As we shall estimate fit by percentages⁴ and as the departures from the Malthusian law are not great the first step is to plot that law, $-\log_{10} P = -0.4343nt$ $\log_{10} B -$ a straight line when P is plotted against t on semi-log paper. The closest fit for New York City and Environs is $\log P = 1.141 + 0.1297t$, where t is measured in decades from the middle year 1855, and the sum of the squares of the residuals is $\sigma^2 = 0.0016$ corresponding to $\sigma = 0.04$ and a "probable error" of 62/3%. The fit is very good,¹³ it is better than that given by Pearl and Reed,³ for which $\sigma^2 = 0.0020$. If we neglect d but now introduce A , we fit¹⁴ $\log (1/P - A) = -0.4343nt + \log B$. For $A = 0.029$ which corresponds to a limiting population of 34.5 millions, we find $\log (1/P - 0.029) = -0.182 - 0.1425t$ with $\sigma^2 = 0.0020$, worse than with $A = 0$. The best fit is very near to $A = 0$ and appears even to be clearly on the negative side (A about -0.01), which would check with our difficulty

in finding an asymptote for the 5-constant solution.^{12,13} Next placing A again equal to 0, we may vary d . For $d = 50$ we find $-\log (P - 50) = 1.110 + 0.1369t$ with $\sigma^2 = 0.0015$ which is better.

There is, however, another method of attack. We have the Pearl-Reed estimates of 197 millions for the whole United States.¹ If Greater New York should maintain its relative position with respect to the rest of the country, its limiting population would be around 17 millions. This would not be in conflict with our fitted curve giving 22 millions for the three states, New York, New Jersey and Connecticut, particularly as the indications are that the states may be entering on a period of somewhat accelerated growth. Let us therefore assume for the City a limiting figure of $A = 0.06$ corresponding to $16\frac{2}{3}$ millions. Let us choose $d = 120$. Then $\log [1/(P - 120) - 0.06] = -0.1352 - 0.1720t$, where t is measured in decades from 1855, and $\sigma^2 = 0.0013$. This is the best fit yet.¹⁵ How much better can be done by suitable variations of d and A we cannot say; the process is tedious. We believe, however, that it is safe to say that the Pearl-Reed method applied with the least squares criterion based on percentages will give a solution for the limiting population much lower than that obtained by them, and correspondingly better suited to our present conceptions of what is economically reasonable and logically not too inconsistent. Such a result is fortunate; for their curves are the best we have, even if, as applied to New York City, there seems to be a great deal of indetermination or divergence in the forecasts.¹⁶

¹ Pearl, R. and Reed, L. J., these PROCEEDINGS, 6, 275 (1920). Reprinted in Pearl's *Studies in Human Biology*, Chap. xxiv.

² Pearl, R., *Proc. Amer. Phil. Soc., Philadelphia*, 63, 10-17 (1924), and studies in *Human Biology*, Chap. xxv.

³ Predicted Growth of Population of New York and Its Environs, 130 East 22nd St., New York City, 1923. Not reprinted in Pearl's *Studies in Human Biology*.

⁴ The predictions of N. P. Lewis, appended to those of Pearl and Reed exceed their figures by 25% in the year 2000, only 75 years hence.

⁵ The mean percentage error is about 6% as against about 8% for the Pearl-Reed fit to New York and Environs. Neither their curve nor ours is a least square solution based either upon the percentage or upon the actual errors. When the populations vary 20 to 40 fold as here there may be considerable difference between solutions based upon different least square criteria. The authors seem now to prefer criteria based on percentages, *Studies in Human Biology*, p. 582, although apparently they used actual differences in the New York paper. Furthermore, as their populations are additive at the asymptotes, it is probable that this condition of additivity was in some way especially introduced as limitation to the least squares fitting of the four individual populations.

⁶ *Studies in Human Biology*, pp. 587, 606.

⁷ The departures from both curves are in fact systematic. In the states the enumerated populations are too high from 1800 to 1840 and too low from 1870 to 1900; in the City the concavity of the curve from 1790 to 1860 is uniformly wrong, and also from 1870 to 1910. The curve $y = \log P = -\log (A + Be^{-nt})$ has the sign of its second derivative d^2y/dt^2 determined by $-Ae^{-nt}$ and is concave down when $A > 0$, concave, up when $A < 0$. Negative values of A mean that the population is growing on a curve which

becomes infinite in finite time. For these three states the concavity is down from 1790 to 1850, but up from 1860 to 1910; what the process of fitting our curve really does is to average in some obscure fashion, the different concavities so that the net result is a downward concavity. It is worth noting that although the concavity in the data for the City is chiefly upward, both early and late, the fit given by Pearl and Reed actually gives a small net downward concavity and hence a finite though very large limiting population. If we should weight heavily the later populations in the three states we might raise our limiting population of 22 millions almost if not quite indefinitely but the fit in the earlier period would become very bad when judged on a percentage basis.

⁸ We refer to the parabolic and logarithmic parabolic curves mentioned by Pearl and Reed, these *PROCEEDINGS*, 6, 1920, p. 275. The Malthusian curve $P = ce^{kt}$ is of course non-additive.

⁹ Any of the Pearl-Reed curves fitted to actual enumerations may bring about such a determination of their constants as to make the population infinite in finite time instead of infinite only at infinite time as was the case with the earlier types of additive curves (or the Malthusian curve) that were used. See note 6 and E. B. Wilson, *Science* 61, 1925 (87).

¹⁰ A simple fictitious example may show this. Suppose that at three equidistant censuses, not necessarily consecutive, the populations of one region are in millions or other units respectively 1, 3, 8 and of another are 1, 4, 9, so that the population of the total region is 2, 7, 17. Then the limiting populations are 33, 12.57 and 30.33 which makes the population of the whole less than that of one part although that of the other part is considerable. The summation of autocatalytic curves has been carefully studied at the Harvard Medical School because of its importance for the theory of the protein titration curve, see E. J. Cohn, *Proc. 11th Internat. Physiol. Congress*, Edinburgh (1923), p. 91.

¹¹ *Studies in Human Biology*, pp. 572-577, 607, 625.

¹² The same appears to be true if the data for 1800, 1830, 1860, 1890, 1920 are used. An equation of effectively odd degree like their xxvii must have at least one real root and their five constant equations can be fitted to either set of data, but the solution gives an infinite population in finite time.

¹³ The Malthusian curve does in fact fit so very well that one may seriously question whether there is any statistical justification for seeking improvement. If we try a parabola $\log P = a + bt + ct^2$ we find $c = +0.00056$, a net upward concavity (see notes 9 and 12), but only equal to its probable error.

¹⁴ The weights for the percentages are no longer quite equal because $\log (1/P - A)$ replaces $\log 1/P$ and the quantity of equal weight is $\log [(1 - AP)/P]$, but the difference is not great.

¹⁵ Whether this last fit is significantly better than any of the others may be doubted. With $\sigma = 0.04$ and 14 data, the standard deviation of σ is by the usual formula $\sigma/\sqrt{2N}$ $0.04/\sqrt{28} = 0.008$ which covers the entire range from 0.036 just found to the Pearl-Reed value 0.045 with considerable margin to spare. The values for the census 1790-1920 for our curve and for the Pearl-Reed curve are given below, and the forecasts from 1930-2000, including that of Lewis.

	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920
Census	225	290	359	428	560	764	1163	1835	2375	3026	3966	5385	7467	8979
W and L	223	273	347	456	613	839	1169	1633	2271	3141	4235	5584	7113	8755
P and R	184	254	350	482	662	908	1242	1692	2295	3095	4131	5460	7117	9122
	1930	1940	1950	1960	1970	1980	1990	2000	Asymptote					
W and L	10370	11840	13110	14130	14890	15460	15870	16160	16667					
P and R	11460	14070	16840	19650	22340	24810	26960	28760	34900					
Lewis	11000	13800	16600	20000	24000	28000	32000	36000	...					

We believe that the Pearl-Reed values are very near to those obtained by basing the least squares criterion on actual instead of percentage differences. If so, the system of weighing determines almost completely the solution obtained.

¹⁶ If it were not for the indeterminateness we should be interested in the fact that for $d = 120$ and $A = 0.06$, the logarithmic errors are not only small but non-systematic and in fact oscillating with an amplitude of about 0.045 and period of about 50 years. A great commercial city like New York might easily show an oscillation of about 10% in population about any law of growth during such major economic cycles as have occurred during the past century with about a fifty-year period (prices high in 1815, low in 1845, high in 1865, low in 1897, high in 1919). The population of the city appears to lag behind the growth curve in extended periods of liquidation and to shoot ahead of it in periods of inflation; but one should not generalize on such meager and speculative material.

ON THE INTERSECTION INVARIANTS OF A MANIFOLD

By J. W. ALEXANDER

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated January 10, 1925

Consider a manifold M that can be subdivided into a finite number of n -cells, bounded and separated from one another by a finite number of cells of lower dimensionalities. If we denote the sensed i -cells of any subdivision S_0 of M by

$$[i]_1, [i]_2, \dots, [i]_{a_i} \quad (1)$$

respectively, the relation of an i -cell $[i]_s$ to the $(i - 1)$ -cell on its boundary may be indicated by writing

$$[i]_s \equiv [i-1]_s^t [i-1]_t, \quad (2)$$

(summed for t), where $[i-1]_s^t$ is a coefficient which vanishes if $[i-1]_t$ is not on the boundary of $[i]_s$, but is otherwise equal to $+1$ or -1 according as $[i-1]_t$ is positively or negatively incident to $[i]_s$.

An i -chain is any linear combination $\lambda^s [i]_s$ of i -cells $[i]_s$, where the coefficients λ^s are integers. Its boundary is the right-hand member of the congruence

$$\lambda^s [i]_s \equiv \lambda^s [i-1]_s^t [i-1]_t \quad (3)$$

derived by combining congruences of the type (2). If the boundary of an i -chain vanishes, the i -chain is *closed*. If an i -chain $(i)_0$ is the boundary of an $(i + 1)$ -chain, the i -chain is *bounding*, or *homologous to zero*: A set $(i)_0 \sim 0$

of closed i -chains is *linearly independent with respect to bounding* if no linear combination of them is bounding. We make the convention (different