V. ESTIMATES DERIVED BY EXTRAPOLATION OF CENSUS RESULTS

A. METHODS REQUIRING TWO CENSUSES ONLY

1. The utilization of the results of two censuses

Estimates based on the results of two or more censuses are more reliable, as a rule, than those based on one census only. As already pointed out, the census data are likely to improve in quality as more experience in census taking is gained, and the availability of a series of census results makes possible a better appraisal of the accuracy of enumeration than is possible with a single census. A comparison of the results of two or more censuses provides, furthermore, a means of estimating the increase of population since the last census.

Extrapolation — that is, the use of the assumption that the mode of increase observed between two dates in the past has continued since the time of the last enumeration — often gives better results than any other method which can be applied in a country where current data on population change are lacking or of poor quality. Extrapolation, however, must not be applied mechanically wherever the results of two or more censuses are available. There may be good reasons to believe that the rate of increase since the last census has differed considerably from the rate during the inter-censal interval, either because of unusual events in one or both of these periods, or because of a general tendency towards acceleration or deceleration in these rates. In some of these cases, a modification of the extrapolation is still preferable to other less well-founded assumptions, though less reliable than time adjustments by means of adequate statistics of births, deaths and migration. Extrapolation depends, strictly speaking, on an assumption that the population change in recent times has not deviated from past trends and that it proceeds, year by year, according to a constant progression. Such precise conformity of biological phenomena to a mathematical formula never occurs in reality. Even though deviations from previously established trends are not always large, new factors of change arise constantly.

In some countries, existing statistics of births, deaths and migration are too inadequate to reveal the true changes in population size. Inadequate migration statistics are almost the rule, but so long as migatory movements are not very large this defect in the statistics is not of major importance. However, if births, deaths or large migratory movements are inadequately recorded, extrapolation may lead to better current estimates than the use of such defective data.

If the results of two censuses only are available, there is a choice between two simple methods of extrapolation: the use of an arithmetic, or a geometric rate of increase. The choice of method should depend on a knowledge of the given situation. For special purposes, more elaborate methods can also be used.[1] The greater variety of methods which can be used if three or more censuses have been taken are discussed in part B of this chapter. The selection of the most suitable method of extrapolation in each case requires good judgment, and no general rule can be stated as to the method which will give the most reliable results in any particular instance.

2. Extrapolation by means of arithmetic rates of increase

The simplest method of extrapolation is to compute the average annual number by which the population has increased from one census to the next, and to add an equal number for every year which has elapsed since the last census.

Example. In Switzerland, 156,715 persons were enumerated in 1936, and 185,215 in 1946. Current estimates are made in the following manner:

The increase observed in the census interval of ten years was 28,500 persons, giving an average annual increase of 2,850. This figure is added to the 1946 census total for each year which has elapsed since that date. The estimate for 1950 was, therefore, 185,215 plus four times 2,850, i.e., 196,615 or, as a rounded figure, 197,000.

It should be noted that a constant numerical increase in a growing population really implies a diminishing rate of growth relative to population size. The use of an arithmetic rate is often justified for some of the following reasons:

1. It is easy to compute an arithmetic rate. If it is doubtful that very reliable estimates can be obtained by any more complicated method of extrapolation, there can be no objection to using the simplest one.

2. An arithmetic rate is less likely to give absurd results than a constant geometric rate of increase if the rate is high and the period of time is long. A population of 10 million increasing at a geometric rate of 2 per cent per annum would attain 79,900,000 within 100 years and 638,470,000 within 200 years. Except in an initially very sparsely settled country, it would appear

[1] For a description of several methods and bibliographic reference pertaining to many other methods of extrapolation, the reader is referred to R. H. Wolfenden, *Population Statistics and their Compilation* (Actuarial Society of America, New York, 1925).

inconceivable that population growth could continue at such a rate for a very long time. An arithmetic rate in such a case gives much more conservative results. Thus, an initial population of 10 million, increasing by 200,000 per annum, would attain 30 million after 100 years and 50 million after 200 years.

3. There are grounds for supposing in some instances that the rate of population growth tends to slacken as time passes. For example, if the birth rate is believed to be declining and the death rate to be nearly constant or declining more slowly, a gradually decelerating rate of growth such as is implied by a constant numerical addition to the population is more reasonable as an assumption than a constant geometric rate. Likewise, if immigration is the source of a major part of the increase and if there is no reason to suppose that the annual number of immigrants increases as the population grows larger, an arithmetic rate may be more reasonable than a geometric one.

4. There is often good reason to assume that enumeration in the earlier census was less accurate than it was in the later one. More often than not a first census results in a considerable under-enumeration which is partially eliminated in the second census. In extrapolating the apparent increase between a first and a second census it may be preferable to be conservative and to use an arithmetic rather than a geometric rate.

5. An advantage of the use of arithmetic rates, which will be discussed again in section 4, arises from the fact that the sum of arithmetically extrapolated sub-totals is always equal to the arithmetically extrapolated total, whereas extrapolation of sub-totals and totals by any other method leads to results which are more or less inconsistent and require further adjustment.

3. Extrapolation by means of geometric rates of increase

Geometric extrapolation corresponds to the assumption that a population increases constantly by numbers proportionate to its changing size. In this case, the computation is carried out in the same manner as a computation of compound interest. A population increasing by a geometric rate obeys the formula

$$Pn = Po\ (1 + r)^{t} \ \ldots\ldots\ldots\ldots (1)$$

where Po is the population at the beginning of a period,

t is the period of time, in years,
r is the annual rate of increase, and
Pn is the population at the end of the period.

If the rate of increase is to be determined from a comparison of census results, it can be found by the formula

$$(1 + r) = \sqrt[t_1]{\frac{P_2}{P_1}} \ \ldots\ldots\ldots\ldots (2)$$

where P_1 and P_2 are the populations according to the first and the second census respectively, and t_1 is the inter-censal time interval, in years.

Thus, for a geometric extrapolation, from the date of the last census onward, of the rate of increase indicated by a comparison of the last census figure with an earlier census figure, the formula can be written:

$$Pn = P_2 \left(\sqrt[t_1]{\frac{P_2}{P_1}} \right)^{t_2} \ldots\ldots\ldots\ldots (3)$$

where t_2 is the post-censal time interval (up to current date).

The application of this formula requires calculating aids (logarithms, a slide-rule, or a compound-interest table). Actually, with a geometric rate of increase, the logarithms of the population increase at an arithmetic rate.

Example. The population of Spain amounted to 23,563,867 according to the census of 31 December 1930, and to 25,877,971 according to the census of 31 December 1940. For the years following 1940, current estimates were made by means of a geometric extrapolation of these two census results.

Using formula (2), the rate of increase can be determined as follows:

$$(1 + r) = \sqrt[10]{\frac{25{,}877{,}971}{23{,}563{,}867}} = 1.0094;\ r = 0.0094.$$

Substituting this rate of increase in formula (1), the following estimate can be obtained for midyear 1950, i.e., after a lapse of 9½ years:

$$P_{1950} = 25{,}877{,}971\ (1.0094)^{9.5} = 28{,}286{,}533,$$

which may be rounded to 28,287,000. (This estimate may have to be revised in view of the fact that the preliminary result of the census taken on 31 December 1950 totalled 27,861,000).

The same result can also be obtained more directly by formula (3):

$$P_{1950} = 25{,}877{,}971 \left(\sqrt[10]{\frac{25{,}877{,}971}{23{,}563{,}867}} \right)^{9.5} = 28{,}286{,}533.$$

Geometric extrapolation always leads to a higher estimate of population size than arithmetic extrapolation whether a population is increasing or declining. Thus, in the above example, if arithmetical extrapolation had been used it would have been found that the population of Spain had increased in the ten years from the end of 1930 to the end of 1940 by 2,314,104, i.e., by 231,410 on an annual average; assuming this arithmetic annual increase to have continued for another 9½ years, up to mid-year 1950, the estimate for the latter date would have amounted to 28,076,366 only as compared with 28,286,533 resulting from geometric extrapolation.

One drawback of geometric extrapolation is that the sum of extrapolated sub-totals is not equal to the extrapolated total.

Example. Let it be assumed that separate geometric extrapolations are made for the urban, the rural and the total population of Canada, up to 1950, using the results of the national census of 1931 and 1941.

In 1931, the urban population was found to be 5,572,058, the rural 4,804,728, and the total 10,376,786. In 1941, the urban population was 6,252,416, the rural 5,254,239, and the total 11,506,655.

Geometric extrapolations up to 1950 would bring the urban population to 6,935,000, the rural population to 5,695,000, and the total population to 12,628,000. The sum of the extrapolated urban and rural populations, however, amounts to 12,630,000, which is slightly more than the extrapolated total population.

While the assumption of a constant geometric rate of increase may be plausible in many circumstances, it cannot be relevant in every situation. Such a rate of increase would occur in a population not affected by migration if the birth and death rates were constant or if both birth and death rates were rising or falling by equal amounts each year. In a population being increased by immigration or depleted by emigration, the rate of growth would be constant only if the net rate of immigration or emigration as well as the birth and death rates were constant in proportion to the population, or if changes in the birth and death rates were exactly counterbalanced by changes in migration.

No country's population can be expected to grow for an indefinite period of time at a constant rate for eventually it would become so large that further increases would hardly be possible. Many countries have experienced in recent times a drop in death rates followed, a generation or so afterwards by a drop in birth rates; as a result, population increased rapidly for a time, but at a slackening pace subsequently; a strictly parallel movement of birth and death rates has hardly been observed over any long time period. It is true that during the period of rapid population growth many countries of Europe sent large numbers of migrants to other continents; this did, however, not entail any large migrations into Europe during the subsequent period of slackening growth.

It is, therefore, desirable to confine geometric extrapolation to relatively short time intervals. For short periods it may be plausible to assume that a given population increases approximately according to a geometric rate and that one of the above-stated combinations of conditions is satisfied to some degree of approximation. If several decades have elapsed since the taking of the last census geometric extrapolation becomes increasingly unreliable and may lead to a cumulative exaggeration of the estimated population, especially if the projected rate of increase is high. It may then be preferable to use arithmetic extrapolation which gives more conservative results. Geometric extrapolation is hazardous also if the accuracy of the basic census figures is subject to considerable doubt; the accumulation of increases which are apparent (only because of varying completeness of enumeration) may lead to greater and greater error in the extrapolation as the years pass.

In some populations growth is for a time accelerated because the death rate is declining while the birth rate remains nearly constant. In such a situation even the geometric rate of increase will underestimate the growth of population and an arithmetic rate is definitely inappropriate. In this event the use of a harmonic rate of increase may be suggested (see section 4, below).

Geometric rates are preferable to arithmetic rates for the extrapolation of decreases in population over a series of years, for an arithmetic extrapolation of a negative rate of increase would eventually result in a negative population estimate, which is obviously absurd. Geometric extrapolation, on the other hand, would show a diminishing numerical decline and would always result in positive population estimates no matter how long the time period. In a decreasing population geometric extrapolation is the more conservative of the two methods. For the same reason geometric extrapolation is probably preferable in countries with a large emigration.

4. Modified methods of extrapolation

As already pointed out, the computation of a geometric rate of increase cannot be carried out without a logarithmic table or other calculating aids. If it is desired to confine the computation to an arithmetical operation, but to obtain a result which approximates that of a geometric rate, the following method can be used.

Instead of keeping the annual increase at a constant figure, one may express the average annual numerical increase between the two census dates as a percentage of the arithmetic average of the totals from the two censuses. This procedure yields a simple approximation to the geometric rate for the purpose of arithmetical computations. This rate can then be applied, year by year, to estimates of a changing population.

Example. Using the data for Spain (see example on p. 29 of this chapter), extrapolation, by arithmetical approximation to a geometric rate of increase, can be performed as follows:

The arithmetic mean of the census results for the end of 1930 and the end of 1940 amounts to 24,720,919. Since average annual increase in this period amounts to 231,410, a rate of 231,410 in 24,720,919, i.e., of 0.936 per cent, can be computed, this being an approximation of the geometric rate. Applying this rate to each successive figure, we obtain the following series:

Date	*Period (years)*	*Population*	*Increase*
31 December 1940	½	25,877,971	121,117
1 July 1941	1	25,999,088	243,351
1 July 1942	1	26,242,439	245,629
1 July 1943	1	26,488,068	247,928
1 July 1944	1	26,735,996	250,249
1 July 1945	1	26,986,245	252,591
1 July 1946	1	27,238,836	254,956
1 July 1947	1	27,493,792	257,342
1 July 1948	1	27,751,134	259,751
1 July 1949	1	28,010,885	262,182
1 July 1950		28,273,067	

The result, 28,273,000, is a close approximation to the result of the actual geometric extrapolation, which was 28,287,000, and may be compared with 28,076,000, the result of arithmetic extrapolation.

This method of extrapolation always results in figures somewhat higher than those obtained by arithmetic extrapolation and differing slightly from those obtained by geometric extrapolation. Like geometric extrapolation, it has the disadvantage that sub-totals do not add up to totals extrapolated by the same method. This method is to be recommended only where it is desired to compute, by approximation, a geometric rate of increase without the use of computing aids.

In some situations the two alternatives of arithmetic and geometric extrapolation may seem to offer too narrow a choice. This is particularly so in the case, already referred to, where population increase is temporarily accelerated. If one wishes to assume an accelerating rate of increase relative to population size, the extrapolation can be carried out with respect to the reciprocals of the numbers involved. A rate of growth accelerating in this manner may be described as a harmonic rate.

Example. Employing again the data for Spain (see example on p. 29 of this chapter), harmonic extrapolation can be carried out as follows:

The reciprocal of the 1930 census figure multiplied by 100 million is 4.243786; and 100 million times the reciprocal of the 1940 census figure is 3.864291. The reciprocal value, therefore, has declined by 0.379495 in ten years, or by 0.03795 per annum. To arrive at an estimate for mid-year 1950, this decline must be extrapolated for 9½ years beyond the value for the end of 1940, and a reciprocal value of 3.864291 minus 0.36052, i.e., 3.503771, is obtained, which corresponds to a figure of 28,540,700 as the estimate for 1950, on the assumption of an accelerated rate of increase. This figure is markedly higher than the result of geometric extrapolation.

If there is reason to believe that the rate of population growth is either accelerating or slowing down, a reasonable assumption as to the rate of acceleration or deceleration can also be made and a formula for modified extrapolation with a changing arithmetic or geometric rate can be derived.

Various methods have been devised for reconciling the sum of geometrically extrapolated sub-totals with a geometrically extrapolated total. In most of these methods the geometric extrapolation for the total is retained and adjustments are applied to sub-totals to make them agree with the total. One method consists in extrapolating sub-totals by means of an arithmetic progression and then multiplying each value by the ratio of the total population derived from geometric extrapolation to the total derived from arithmetic extrapolation. This method has the drawback that the sub-total for an area which had a stationary population in the inter-censal interval would show a decline in the post-censal period if the total population were increasing. Another method which was developed by A. C. Waters consists in an arithmetic extrapolation of the ratio of the sub-total to the total. The disadvantage of this method is that the sub-total for an area with stationary inter-censal population would show a post-censal increase if the total population were increasing. Other methods have also been developed but they require more complex mathematical operations.

B. METHODS REQUIRING THREE OR MORE PREVIOUS CENSUSES

5. Extrapolation by means of parabolas

The question as to whether the rate of population growth tends to be constant or to accelerate or slow down cannot be answered with much assurance by comparing successive census totals unless the results of at least three censuses are available. But even the results of three censuses are insufficient for determining whether an observed acceleration or deceleration of increase appears to be temporary or a long-term trend, whether it is fairly recent and may still gather momentum, or whether it has apparently run its course. To answer such questions with any assurance a minimum of four census totals is required.

It is, of course, not necessary to employ the results of more than two censuses in an extrapolation even though more are available. If the results of a series of censuses suggest that population growth has proceeded at a fairly even pace, and if there is reason to suppose that it may continue to do so, arithmetic or geometric extrapolation is quite sufficient for most purposes, at least if the post-censal period is not too long. Similarly, if one or several of the census results are of doubtful accuracy, those results may be selected for purposes of extrapolation which are believed to be most nearly comparable. Extrapolation by methods more refined than arithmetic and geometric rates is hardly justified if a basic figure is unreliable.

The most widely used method of extrapolation which employs the results of three or more enumerations is extrapolation by parabolas of the second or the third degree. A parabola of the second degree can be computed from the results of three censuses; this type of curve is sensitive not only to average rates of growth but also to the observed acceleration or deceleration in these rates. A parabola of the third degree, which can be computed from the results of four censuses can take account not merely of the acceleration or deceleration in rates of growth, but also of a changing momentum in the acceleration or deceleration.

If the results of three censuses are plotted on a chart where the horizontal axis measures intervals of time and the vertical axis the population enumerated at each of the three dates, these points can be connected by a parabolic curve of the second degree which, if continued beyond the date of the last census, indicates the population which may be estimated for each subsequent date. This parabolic curve is defined by the formula

$$Px = a + bx + cx^2,$$

where x is the time interval, in years, measured from any fixed date, such as perhaps one of the census dates.

Px is the population size to be expected at a time of x years after the fixed date,

any fixed date, such as perhaps one of the census dates.

solving the above formula for each of the three census dates.

Example. Let it be assumed that, according to a census of 1936, the population was 40,000; according to a census of 1942, it was 44,000, and according to a census of 1946, it was 47,000. (For the sake of simplicity, exact calendar years and round population figures are used in this example.) How is the population in 1950 to be estimated, using parabolic extrapolation?

Taking the year 1936 as the fixed date, $x = 6$ for 1942, 10 for 1946, and 14 for 1950. The equations can be set up as follows:

(1) for 1936: $40{,}000 = a + 0.b + 0.c$ .... $(x = 0)$
(2) for 1942: $44{,}000 = a + 6b + 36c$ .... $(x = 6)$
(3) for 1946: $47{,}000 = a + 10b + 100c$ .... $(x = 10)$

The solution of these simultaneous equations gives the following values:

$$a = 40{,}000$$
$$b = 616.7$$
$$c = 8.33$$

The formula, $P_x = a + bx + cx^2$ may now be expressed as:

$$P_x = 40{,}000 + 616.7x + 8.33x^2.$$

Applying this formula for 1950 ($x = 14$), we obtain:

$$P_{14} = 50{,}267.$$

Computations are simpler if the time intervals are equal. In that event it is convenient to take the date of the second of the three censuses as the fixed date.

Extrapolation of this type has been employed for current population estimates of the Dominican Republic in the years 1920-49. Current estimates of Peru in the years after 1940, on the other hand, were made by the use of a parabola of the third degree, based on the censuses of 1836, 1850, 1876 and 1940.

A parabolic curve of the third degree is defined by the formula

$$Px = a + bx + cx^2 + dx^3.$$

The computation of these constants requires four separate equations, which must be solved for the four census dates.

Example. Censuses were taken in a country in 1910, 1920, 1930, and 1940, resulting in population totals of 250,000, 300,000, 320,000, and 360,000, respectively. How is the 1950 population to be estimated, using extrapolation by a parabola of the third degree?

(1) $250{,}000 = a + 0b + 0c + 0d = a$
(2) $300{,}000 = a + 10b + 100c + 1{,}000d$
(3) $320{,}000 = a + 20b + 400c + 8{,}000d$
(4) $360{,}000 = a + 30b + 900c + 27{,}000d$

Solving these four simultaneous equations, we obtain the following values:

$$a = 250{,}000$$
$$b = 8{,}166.7$$
$$c = -400.$$
$$d = 8.333$$

To obtain the estimate for 1950, we have to substitute a, b, c, and d, for $x = 40$, in the general formula of the third-degree parabola.

$$P_{40} = a + 40b + 1{,}600c + 64{,}000d, \text{ i.e., } 470{,}000.$$

(If a second-degree parabola had been used in conjunction with the census figures of 1920, 1930 and 1940, the population estimate for 1950 would have amounted to 420,000. The difference in the results is due to the fact that, in the present example, the second-degree parabola would have expressed some acceleration of population growth, whereas the third-degree parabola also expresses an increasing momentum of acceleration.)

The selection of a second-degree or third-degree parabola for extrapolation presupposes, of course, that the form of growth expressed by such a curve appears plausible in view of what is known about the trends of births, deaths and migration. It may seem advisable to use a parabola of the second degree if, for instance, the birth rate is believed to be rising while the death rate falls, so that the rate of population growth is accelerating; or if there is reason to suppose that the rate of growth is slackening because of a falling birth rate and a more nearly constant death rate. A parabola of the third degree can take into account changes in the speed at which the gap between birth and death rates is widening or narrowing; for instance, when death rates fall rapidly at first, but less rapidly as time progresses. An increasing or decreasing volume of migration can also lend support to the assumption of a parabola.

Compared to arithmetic and geometric rates, parabolas have much more flexibility. It would, however, be fallacious to expect that population growth will continue to conform to some parabolic trend, merely because it has done so in the past. Too much refinement may also lead to an illusion of increased accuracy in the estimates although estimates must always be less reliable than the basic data.

A very important defect of parabolas, if they are extrapolated over a considerable length of time, is that they eventually move more and more rapidly either in an upward or a downward direction. This tendency is more pronounced in third-degree curves than in those of the second degree, owing to their greater flexibility. Although, for a time, population growth may indeed be accelerating, there is an upper limit to the rate of growth which can ultimately be attained. There is also a lower limit to population size, which can never become negative, whereas some parabolas may eventually fall below zero.

In most situations, a parabola of the second degree should be preferred to one of the third degree. The additional refinement of the latter is usually not warranted by the available data, nor can it be very realistic to expect population growth to conform to a function of such complexity. In the particular case, however, the use of a third-degree parabola is sometimes suggested when it appears that it leads to an extrapolation with less exaggeration than a second-degree parabola.

6. Parabolic extrapolation of transformed data

The tendency of parabolas after several years of projection to show a steeper and steeper rise or fall has already been mentioned. In many cases, this defect can be modified by applying parabolic extrapolation to the logarithms of the figures instead of to the figures themselves. The extrapolation of logarithms implies a projection of changing rates of growth rather than of changing absolute figures.

Example. Using the data in example on pp. 31-32, a second-degree parabola can be computed on logarithms as follows:

Date	*Number*	*Log.*	*Equation*
1936	40,000	4.60206	$\log. P_0 = a + 0b + 0c = 4.60206$
1942	44,000	4.64345	$\log. P_6 = a + 6b + 36c = 4.64345$
1946	47,000	4.67210	$\log. P_{10} = a + 10b + 100c = 4.67210$

Solving these equations, we obtain:

$$a = 4.60206$$
$$b = .006742$$
$$c = .000026$$

To obtain the logarithm for the 1950 estimate, these values must be substituted in the equation.

$$\log. P_{14} = a + 14b + 196c$$

The logarithm for the 1950 estimate, therefore, is 4.701544, and the 1950 estimate is 50,297. This result of logarithmic extrapolation compares with 50,267, resulting from direct extrapolation.

Another transformation of data which can sometimes be used with advantage consists in using reciprocals. As has already been mentioned, extrapolation of reciprocals implies an assumption of accelerating growth, which is realistic in some instances. Parabolic extrapolation of reciprocals may be pertinent where estimates derived by direct parabolic extrapolation tend to be too conservative or result in rapidly decreasing figures.

7. Extrapolation by means of growth curves

For purposes of estimating the population at future dates or at dates in a distant past when no adequate information was given, highly elaborate methods of extrapolation are often employed. Most important among these is the logistic curve of growth, first discovered by P. F. Verhulst and later developed by R. Pearl and L. I. Reed.[2]

The logistic curve shows a long-term growth cycle, usually extending over a century or longer, which contains the following phases: (1) a gradual transition from almost stationary conditions to a noticeable increase in the population, (2) an acceleration in the rate of increase until the rate approaches a maximum, (3) a slowing down of the rate of increase, and (4) a gradual transition towards almost stationary conditions. It has the advantage that it can fit any set of data showing at least one of these phases of growth.

It has been demonstrated that population growth over periods of a hundred years or more in many countries has approximated fairly closely a logistic curve of growth. It has also been shown that logistic curves possess a certain predictive value and that future estimates made by means of logistic extrapolations have in many cases been approximately confirmed by actual observation as censuses were taken subsequently.

Studies of these kinds have led some scientists to the belief that the logistic curve expresses some universal law of population growth. Experiments were made with cultures of yeast-cells or with a colony of fruit-flies in a bottle and it was found that biological organisms, multiplying within a limited or closed environment, showed changing phases of growth closely resembling those of the logistic. It does not follow, however, that the same applies to human populations since human beings, as distinguished from flies in a bottle, have both the capacity to change their environment and to control their rates of reproduction.

The Gompertz curve has also sometimes been used as a curve of population growth.[3] Its main difference from the logistic curve consists in the fact that it expresses a more rapid acceleration in the early phases of growth and a more gradual slowing down in the late phases.

Growth curves are used only very seldom for purposes of current estimates. A logistic curve was computed by the Banco Nacional of Cuba,[4] to estimate the population of Cuba for the years 1937-49 because estimates made at various times during that period by the Cuban census office, using defective statistics of births and deaths, were found inconsistent.

Although a simple logistic or a Gompertz curve can be fitted to the results of three censuses, and a modified logistic curve to the results of four censuses — provided these show a somewhat regular procession of growth — this is rarely done. Growth curves are usually computed from a long series of census figures in such a manner that they do not coincide exactly with any one of the census results, but approximate each and all of them. The reason for this practice is that individual census results are affected not only by the long-term trend in population growth, but also by temporary departures from this trend, whereas growth curves are intended to portray the long-term trend only. This makes it inadvisable to employ growth curves in an extrapolation of the results of three or four censuses only. Any particular census result may deviate more or less from the over-all trend and an extrapolation of such temporary deviations from the trend would lead to unrealistic results and defeat the purposes for which these elaborate curves are intended. Nearly all countries which have a long series of census results also possess adequate systems of birth and death registration, making extrapolation unnecessary for the purpose of current estimates.

C. GENERAL OBSERVATIONS PERTAINING TO CURRENT ESTIMATES DERIVED BY EXTRAPOLATION

8. Elements of reliability in estimates derived by extrapolation

The following are the main factors affecting the reliability or reasonableness of extrapolation:

1. The accuracy of the census data;

2. The relative accuracy, or comparability, of the data from the different censuses;

3. The length of inter-censal intervals and of the post-censal period; and

4. The relevancy of the method of extrapolation selected.

[2] Hagood, M. J. *Statistics for Sociologists.* Henry Holt & Company, New York, 1951, pp. 272-282.

[3] Croxton, F. E., and Cowden, D. J. *Applied General Statistics.* New York, Prentice-Hall, Inc., 1950, p. 448.

[4] Cuba. *Banco Nacional de Cuba.* Memoria 1949-1950, La Habana, Cuba, p. 155.

The accuracy of the census results is important because a current estimate, made by extrapolation, cannot be more accurate than the base data. If the probable direction and possible extent of errors in the census figures (due to over- or under-enumeration) are known, the census data should be "corrected" before they are used in the extrapolation, as this will reduce the margin of error in the resulting estimates.

Differences in the accuracy of various census results can affect extrapolation even to a greater extent than constant errors running through a series of census enumerations. For example, if in two successive censuses there was under-enumeration of about the same amount, a comparison of the results leads to an estimated rate of population increase which is fairly realistic. Although the results of an extrapolation of these census totals are likely to be underestimates, this component of error is not greatly increased by extrapolation. However, if two successive censuses have differed with regard to the extent of under-enumeration or over-enumeration, or if they have erred in opposite directions, a comparison of the totals can lead to a computed rate of increase which is much too high or too low. Extrapolation of such an erroneous rate of increase will probably lead to even greater error in the resulting estimates. This is also true if one of two censuses was fairly accurate, but the other was not. Thus, a rate which is extrapolated from an earlier census which was notably incomplete and a later census which was more accurate is likely to lead to overestimates of the current population; this error increases every year. The possible errors in the successive censuses should therefore be given careful consideration, and if there is reason to believe that they have varied in accuracy, the totals should be "corrected" before extrapolation.

Consideration should be given also to the comparability of the census results with regard to the definition of the population, the inclusion or exclusion of special categories, and the territory covered. It would clearly be wrong to extrapolate census totals without adjustment if one census was taken within a larger or a smaller territory than the other. Similar errors may result if the censuses used in extrapolation differ in coverage of particular categories of the population.

The reliability of an extrapolation is affected by the length of inter-censal intervals. A long interval can be favourable for two reasons: year-to-year fluctuations in population increase tend to be compensated over a longer period and true trends are reflected with more weight in a long period than incidental differences in completeness of census enumerations. A short interval, on the other hand, has the advantage that it refers to a less remote past and may therefore give a somewhat better basis for estimating trends in the most recent years, and that census results obtained within a relatively short succession are likely to be more nearly comparable.

In some instances it may be desirable to select two or three censuses from a series of censuses in such a manner that extrapolation can be made on the basis of inter-censal intervals of a desirable length. Where population trends are apt to change rapidly, more reliable extrapolations can be obtained if they can be based on rather short intervals. Longer intervals should be selected, if possible, where population trends are fairly constant or subject to gradual changes only.

The reliability of an extrapolation depends directly on, and is in inverse proportion to, the length of the post-censal period, i.e., the period between the last census and the date for which an estimate is desired. As the post-censal period becomes longer, errors may accumulate more and more rapidly, because recent trends deviate more and more from extrapolated trends; this is of necessity true in the case of parabolic extrapolation. In particular, it is undesirable to continue extrapolating over a post-censal period which is as long as, or even longer than, the inter-censal intervals. If a long time has elapsed since the last census was taken, it may be desirable either to base extrapolation on longer inter-censal intervals (provided a sufficiently long series of previous census results is available), or to abandon the method of extrapolation and to make current estimates by other methods. In particular, parabolic extrapolation should not be carried over a long period (in view of the tendency of parabolas to rise or fall more and more steeply after a certain time has passed), unless it is based on very long inter-censal intervals. Thus, with the passage of time, parabolic extrapolation may have to be replaced, first, by extrapolation of arithmetic or geometric rates and, eventually, by different types of estimates, such as those based on the results of more recent non-censal counts, or even conjectures.

A consideration of all the factors discussed in this section (accuracy and comparability of census results, length of inter-censal and post-censal periods) as well as any knowledge regarding probable recent developments affecting the countries' population must influence the choice of the method of extrapolation. A method can be regarded as appropriate if it leads to estimated post-censal increases in the population which do not appear in any way unrealistic. The selection of a method of extrapolation depends, therefore, largely on sound judgment and a knowledge of the situation. If extrapolation is made by an unrealistic method, errors in estimates are likely to increase more and more rapidly in the course of years.

9. Improvement of estimates derived by extrapolation, and standards of comparability

Using the same methods of extrapolation, estimates can be improved by either selecting the most useful censuses if a sufficient number of census results are available, or by "correcting" census results where this seems necessary. Censuses may be so selected that time intervals which are too short are avoided and census figures that are not comparable are omitted. If the last census interval was unusual with respect to rates of increase (e.g., owing to temporary disturbance of normal trends, or owing to a war), an arithmetic or geometric rate of increase observed in some previous census interval may be applied to the total figure obtained at the last census.

Fresh consideration should continuously be given to the selection of the method of extrapolation, particularly if a long time has passed since the last census. In particular, parabolic extrapolation may have to be abandoned after several years and some simpler method substituted. Eventually, any method of extrapolation must be abandoned and estimates have to be formed on some new basis rather than by relying on remote census results.

If it is believed that recent developments have caused the growth of population to depart from previous trends, all knowledge regarding the possible influence of such developments on birth and death rates or on the volume of migration must be taken into account in selecting the most suitable method of extrapolation or in modifying the results of an extrapolation. If unusual events have occurred in a particular year only (e.g., a famine resulting in much excess mortality, the arrival of large numbers of refugees, etc.) the increase or decrease of population will have to be estimated separately for that year by some other means, but extrapolated rates of increase may again be applied for subsequent years.

The best way of improving estimates derived by extrapolation is, of course, by taking a new census. So long as no new census is taken, it is important to indicate, so far as possible, the reliability of estimates. One way of ascertaining approximate margins of error is to perform several independent extrapolations, one by assuming a rate of increase which can be regarded as a maximum, and one by assuming a rate which would seem to be a minimum.

If the results of the last census had to be modified in order to derive comparable figures for purposes of extrapolation, this fact should be indicated. Current population estimates are frequently used in conjunction with the results of the last census, but a comparison of these two figures leads to fallacious conclusions unless it is made clear how the extrapolated figures differ from those of the census. If a census result was increased to allow for an estimated amount of under-enumeration, or for a category of the population which had not been enumerated, this should be explained.

10. Comparative results obtained by various methods of extrapolation

In order to illustrate the results obtained by various methods, estimates of population at the date of the last census have been made, for several countries, by extrapolating in different ways the trends shown by earlier censuses. Comparison between these estimates and the enumerated figures for the latest census dates gives some insight into the types of errors produced by the different methods of extrapolation under various conditions. The countries selected for these illustrative calculations differ greatly as regards population trends, length of inter-censal intervals, and presumable accuracy of census enumeration.

The following types of extrapolation were carried out:

(1) Arithmetic rate of increase between two preceding census dates;

(2) Geometric rate of increase between the two preceding census dates;

(3) Parabola of the second degree fitted to the totals from three preceding censuses;

(4) Parabola of the third degree, fitted to the totals from three preceding censuses;

(5) Second-degree parabola computed on logarithms of the totals from three preceding censuses;

(6) Third-degree parabola computed on logarithms of the totals from three preceding censuses.

For British Honduras in 1946, for example, the following results were obtained:

Population according to the 1946 census: 59,220

Result of extrapolation (1):	60,342,	error:	+ 1.9 per cent
Result of extrapolation (2):	61,865,	error:	+ 4.5 per cent
Result of extrapolation (3):	62,588,	error:	+ 5.7 per cent
Result of extrapolation (4):	70,342,	error:	+ 18.8 per cent
Result of extrapolation (5):	63,280,	error:	+ 6.9 per cent
Result of extrapolation (6):	59,856,	error:	+ 1.1 per cent

The population of British Honduras has been growing from census to census at first at a rising, but more recently at a declining rate. It grew from 37,479 in 1901 to 40,458 in 1911, to 45,317 in 1921, to 51,347 in 1931, to 59,220 in 1946. Hence the tendency of all methods of projection was to exaggerate the result for 1946. In this case, method (6), a third-degree parabola computed on the logarithms of the data furnished the best result but method (4), a third-degree parabola computed on the data themselves, furnished the worst result.

Errors resulting from extrapolations by the same six methods for fifteen countries are shown in table I.

Table I

ERRORS RESULTING FROM EXTRAPOLATION OF PREVIOUS CENSUS RESULTS TO DATE OF LAST CENSUS, BY SIX METHODS OF EXTRAPOLATION[a]

Country and date of last census[b]	*Interval since last previous census (years)*	*Per cent error* Method (1)	Method (2)	Method (3)	Method (4)	Method (5)	Method (6)
Egypt, 1947	10	− 7.9	− 6.9	− 5.7	−12.5	− 5.4	− 2.9
Mauritius, 1944	13	− 1.4	− 1.2	+ 1.9	+ 0.7	− 0.2	+ 0.8
Canada, 1941	10	+ 4.0	+ 6.5	+ 4.1	+ 6.3	+ 3.1	+ 8.9
Alaska, 1950	11	−32.8	−30.3	−24.2	−21.8	−18.6	−26.5
British Honduras, 1946.	15	+ 1.9	+ 4.5	+ 5.7	+18.8	+ 6.9	+ 1.1
Puerto Rico, 1950......	10	− 0.5	+ 1.7	+ 2.7	−16.9	+ 4.1	−20.5
Ceylon, 1946	15	−14.1	−13.2	+ 9.6	+28.2	+11.7	+53.9
Thailand, 1947	10	+ 4.6	+11.1	+12.1	+13.8	+17.6	+14.4
Denmark, 1940	5	− 1.6	− 1.5	− 1.4	− 2.2	− 1.4	− 2.3
France, 1946	10	+ 2.4	+ 2.5	+ 1.4	−12.4	− 6.8	−11.5
Ireland, 1946	10	+ 0.3	+ 0.3	+ 3.3	+ 2.1	+ 9.8	+ 7.1
Portugal, 1940	10	− 1.3	+ 0.01	+ 8.4	+22.5	+12.3	+34.7
Switzerland, 1950	9	− 6.1	− 5.9	− 6.2	− 7.3	− 6.2	− 7.5
New Zealand, 1951	6	− 8.5	− 8.3	−11.5	− 5.5	− 8.6	− 3.7
Western Samoa, 1945..	9	+ 0.8	+ 7.4	+16.1	+31.5	+24.4	+ 0.1
Mean error, irrespective of sign		± 5.9	± 6.8	± 7.6	± 13.5	± 9.1	±13.1

[a] The methods are described on page 35. Data from *Demographic Yearbook*, 1951.
[b] Last census for which final results were available.

Only a few conclusions can be drawn from this comparison. It appears that projections of any kind are liable to err greatly in those countries where population growth has been irregular in the past. It also appears that extrapolations over a time period about equal to preceding census intervals can lead to large errors, particularly when third-degree parabolas are employed. However, parabolas of the higher order may be more reliable than some of the simpler methods if only a short time period is employed for projection from the last census. This seems to be the case in Ireland, where the last time interval was five years, and in New Zealand, where it was six years, while in most other cases the last time interval was ten years or more.

Method (1) (arithmetic rate) furnished the best results for Puerto Rico, Thailand and Ireland, but the poorest result for Alaska; method (2) (geometric rate) furnished the best results for Portugal and Switzerland, and in no case the poorest; method (3) (second-degree parabola) gave the best results for Ceylon, Denmark and France, and the poorest for Mauritius and New Zealand; method (4) (third-degree parabola) gave the poorest results for Egypt, British Honduras, Denmark, France and Western Samoa, and in no case the best; method (5) (second-degree parabola on logarithms) gave the best results for Mauritius, Canada and Alaska, but the poorest for Thailand and France; method (6) (third-degree parabola on logarithms) gave the best results for Egypt, British Honduras, New Zealand and Western Samoa, but the poorest for Canada, Puerto Rico, Ceylon, Portugal and Switzerland.

For Egypt, Alaska and New Zealand, all the methods used result in underestimates because population growth was accelerated in the most recent period. In each of these three cases, a third-degree parabola computed on the reciprocals of the data furnished a better result than any of the above methods, with an error of + 2.8 per cent for Egypt, — 6.3 per cent for Alaska, and — 2.3 per cent for New Zealand. In the cases of Denmark and Switzerland, where the above methods of extrapolation also resulted consistently in underestimates, no advantage was gained by an extrapolation of reciprocals.

The only general conclusion which can be drawn from this comparison is that a method which can give the best result in one situation may give the worst result in another situation. Good judgment must be used in selecting the most suitable method of extrapolation for the purpose of obtaining current population estimates in any given situation.

11. Short-period extrapolations for the purpose of making provisional estimates

In countries where births and deaths are regularly recorded it is sometimes necessary to obtain a preliminary population estimate for immediate purposes before complete statistics of births and deaths for the preceding year or some other brief recent interval have become available. Such a provisional estimate may be made by means of extrapolation.

The length of the period over which extrapolation must be extended in such cases depends on the regularity and promptness with which complete statistics of births and deaths are obtained. Usually, this period is very short, perhaps a quarter or half a year. Much refinement is not necessary because the error resulting from extrapolation over such a short period can hardly be great. The choice of method is therefore not very important and arithmetic rates, being the simplest for computation, are ordinarily satisfactory, unless recent population changes have been of an unusual character.

If there is reason to believe that trends are fairly stable, extrapolation may be based on the changes during a period of a few recent years. If trends have been changing, it is preferable to use some other time interval during which population changes were of the kind which may have occurred during the most recent period. Thus it may be best to base such extrapolations on the last available census result and on the last available population estimate (such as, e.g., the official estimate for the middle, or the end, of the last year for which an estimate has been made by means of vital statistics).

If it is preferred for any reason to use a different method, such as geometric extrapolation, it is better to extrapolate numbers or rates of births, deaths, immigration and emigration, than to extrapolate population size.

A provisional estimate derived in such manner should be indicated as "provisional" in order to avoid confusion when it is replaced by a firmer estimate.