CHAPTER TEN

POPULATION ESTIMATES, PROJECTIONS, AND FORECASTS

Early in Chapter 1, we observed that demography is not only a logically well-organized and highly quantified scientific discipline in its own right, but that it also is one of the most widely applied of all the academic social sciences: in government, business, and other areas. By this point in our discussion, it should be easier to see why this is so. For the methods and theories that help us understand the characteristics of current and past populations, from fertility analysis, to migration studies, to the interpretation and creation of life tables, have turned out to be very useful in many modern occupations and professions. To extend the theme further, in this chapter our attention is turned to a set of techniques that has been developed specifically with such practical applications in view; that is, approaches to anticipating the nature of future populations—or of present populations with past data. These have assumed a unique and indispensable niche in contemporary decision making, especially in the fields of urban and regional planning, public administration, and education, health, and social service management.

Demographers employ several techniques in attempting to understand how populations at a future date (or at a date later than that for which data are available) will resemble or differ from their current state. These are generally classed into three categories, each of which is based on a specific set of assumptions and mathematical formulations: (1) estimation, (2) **projection**, and (3) forecasting. Following a brief introduction, the following sections provide explanations of each type, in the order mentioned, along with extended mathematical derivations and illustrations using census data, current information from other sources, and formal models.[1]

Before we begin, however, it might help to clear away a source of possible misunderstanding. As interesting and useful as projection and the other techniques may be, they do not produce certain answers about tomorrow's populations; demographers do not possess crystal balls or tarot cards that let them "in" on what will happen in the future. Rather, all of these techniques use factual information about the present and the past to help provide educated guesses about what would occur if certain assumptions were to hold. In this way, they do reveal much authentic information about the nature of the present and past data on which they are based, but they do not and cannot produce "facts" about the future. Nathan Keyfitz (1968:27), one of the world's leading experts on population projection, put the matter succinctly. "The object [of projection] is to understand the past rather than to predict the future; apparently the way to think effectively about a set of observed birth and death rates is to ask what it would lead to if continued."

APPROACHES TO ANTICIPATING POPULATION CHARACTERISTICS

In ordinary conversation, we tend to employ certain words that refer to anticipating the future more or less interchangeably. Among the most frequently used are predict, estimate, project, and **forecast**. In scientific applications, however—including population science, each of these has a slightly different meaning which, if not respected, can lead to serious confusion.

Perhaps the most difficult of these four is the first, *predict* and the noun form, **prediction**. In ordinary usage, a prediction is a statement about what we think will happen in the future, based on fact or intuition; for example, "I predict that my car will hold up long enough to get me to work this morning." As we shall see, this is best called a "forecast" in scientific contexts. According to technical usage, a prediction is virtually identical to an explanation (see Rudner 1966). It is the outcome of a strict deductive process that need not even apply to the future at all. In contrast, a population *estimate* is an assessment of a population's size or other characteristics at a present or near-future date, for which we have no immediately current information. Estimates are actually updates of old data, based on the most recent data available. When census data are used, the procedure takes place either at an intercensal date (e.g., 1985, 1995, or 2005) or immediately following the most recent census count.

Estimates are necessary because it is practically impossible to collect data continuously, especially the amount and type gathered by government census operations. Even the kinds of surveys that are regularly conducted by the United States and other national census bureaus, which are very useful in validating estimates, are out of date the moment they are concluded: people are born, and migrate, and die whether or not interviewers are in the field. Yet, a continual demand exists for fresh information about population size, structure, and vital events from national, state, and local governments to allocate financial resources and for other purposes.

The point in time (usually the most recent point) for which authoritative data on demographic characteristics exists, such as Census years 1980, 1990, 2000, is referred to as the "**central date**." The size of a population at midyear of that central date, for example, July 1990, is used as the denominator of crude birth rates, death rates, and many other measures, as we saw in earlier chapters. But if we would like to know something about vital rates at a date between 1990 and 2000—for instance, birth rates in 1996—estimation is necessary so that the appropriate denominator can be determined (recall that information about the vital events in the numerator is continuously registered). To illustrate, the midyear 2010 U.S. population was enumerated at approximately 308,745,538. Based on this and other information, the following set of midyear estimates through 2013 (in thousands) was calculated by the Census Bureau:[2]

2010: 309,326
2011: 311,582
2012: 313,873
2013: 316,128

A *projection* depicts likely population characteristics at a future date based on a set of explicitly stated assumptions about what is expected to occur between the time the projection is made and the date to which it applies. The accuracy of a projection depends upon how closely the assumptions agree with the events that actually occur. Because it is usually difficult to assess future trends, several different scenarios are employed to account for a range of reasonable outcomes. This results not in one single projection but, instead, in a series of projections, of which one might assume slow growth, another moderate growth, and a third rapid growth. The U.S. Census Bureau has used this technique routinely for many years in its household projections, and the results are now regularly updated and posted at the *www.census.gov* web site.

More recently, projections of all the world's national aggregate populations for which information is available have been undertaken by the U.S. Census Bureau, national governments, the United Nations, and private organizations. Table 10.1 contains a set of moderate-growth projections for the total size of the world's population and selected nations to midyear 2025 and 2050, drawn from various sources by the Population Reference Bureau, Inc. The table shows that the world as a whole and most countries will continue to grow fairly rapidly well into the twenty-first century. However, a few countries are likely to experience little or no growth, and some—including Italy and Russia—are expected to experience significant population declines, as they were below ZPG at the time the projections were derived.

With the world's population size approaching 6.9 billion in 2010, the year on which the calculations were based, the total for the year 2025 was projected to be just over 8 billion. Using the 2010 world growth rate of 1.2 percent per year and the model of exponential growth (see Chapter 8), if no changes occurred in birth and death rates, the 2025 total would be slightly over 8 billion.

In the case of the United States' projection to 2025, with a rate of natural increase of 0.6 percent in 2010, the total projected population size in 2025 would be 351 million. In contrast, India's population growth is expected to decelerate over the projection period, from its then-current rate of 1.6 percent per year to 1.5 percent between 2010 and 2025. Nevertheless, with just about 1 billion people at the beginning of the projection period, this figure will probably reach 1.75 billion by the end of the period.

A *forecast* is neither fact nor sheer fiction (see Worrall 2014; Goodman 1983). Rather, it is an assessment of a future state of affairs, including the future state of population characteristics, based on any or all of several sources: projection, scientific theory, intuition, and even sheer guesses. It is the most idiosyncratic and least systematic of the three techniques, and often depends upon the personal viewpoint of the forecaster. Although a forecast is not as reliable as an estimate or a projection, and cannot replace these in attempting to anticipate tomorrow's demographic events, it does have its place in the larger scheme of things because it is so eclectic. We are well aware that the size, growth rate, rate of vital events, age structure, and other aspects of today's populations were affected not only by yesterday's size, growth rate, and so on, but also by a range of other, nondemographic factors. These include population policies and policy shifts, as discussed in Chapter 11, economic conditions, environmental changes, and even natural disasters. The same can be said of the range of possible causes today that will affect tomorrow's populations.

Table 10.1. Projected Population Sizes (in Millions) of the World and Selected Countries to 2025 and 2050

Country	2010 Population Size	2025 Projection	2050 Projection
World	6892	8108.0	9485.0
United States	309.6	351.4	422.6
Canada	34.1	39.7	48.4
Mexico	110.6	123.4	129.0
Haiti	9.8	12.2	15.7
Jamaica	2.7	2.9	2.7
Puerto Rico	4.0	4.1	3.7
Argentina	40.5	46.2	52.4
Brazil	193.3	212.4	215.3
Peru	29.5	34.5	39.8
Australia	22.4	26.9	34.0
New Zealand	4.4	5.0	5.6
Papua New Guinea	6.8	9.1	13.4
Algeria	36.0	43.6	50.4
Egypt	80.4	103.6	137.7
Sudan	43.2	56.7	75.9
Ethiopia	85.0	19.8	173.8
Ghana	24.0	31.8	44.6
Nigeria	158.3	217.4	326.4
Kenya	40.0	51.3	65.2
Tanzania	45.0	67.4	109.5
South Africa	49.9	54.4	57.4
Israel	7.6	9.4	11.4
Saudi	29.2	35.7	49.8
Turkey	73.6	85.0	94.7
China	1338.1	1476.0	1437.0
India	1188.8	1444.5	1748.0
Japan	127.4	119.3	95.2
United Kingdom	62.2	68.6	77.0
France	63.0	66.1	70.0
Germany	81.6	79.7	71.5
Albania	3.2	3.3	2.9
Italy	60.5	61.9	61.7
Spain	47.1	48.4	49.1
Poland	38.2	37.4	31.8
Russia	141.9	140.8	126.7
Ukraine	45.9	41.9	35.3

Source: Haub, Gribble, and Jacobsen (2011).

When looking to the future, a forecaster can take these nondemographic factors into account and, under the right circumstances, provide a useful depiction of the shape of things to come. We have just noted that recent projections assume that India's growth rate will decline over the next few years, largely because of declining fertility rates. But it is within the realm of possibility that the political winds shift in India, and a new government is elected which, like that of Iraq, is strongly pronatalist and does not view the present growth rate as too high. Such a government might take steps to withdraw support from or even close down the country's extensive system of family planning clinics. Now, suppose further that some forecasters with keen political insight sense these changes just beginning to unfold and they include such information in their characterization of the country's future population size. If this were to happen, the old projections would prove to be less accurate than the forecast.[3]

POPULATION ESTIMATION

Several methods of estimation are used by government census offices and academic and private researchers. An estimate is an attempt at arriving at the size of the current population based on data that reflect current or recent conditions. Most of these rely on mathematical methods, including those based on simple growth models and on the fundamental equation of demography: Growth = Natural Increase + Net Migration. We have already employed the simple growth models in other contexts, but they are analyzed here for the first time.

Simple Growth Models

The simple growth model approach seeks to determine the unknown size of a population. This can be (1) at a specific point that lies between two dates for which information is available (e.g., an intercensal period), or (2) that occurs soon after a central date at which size is known (e.g., immediately following a census count). It uses one or more of several mathematical formulas to depict the nature of change and the estimated population sizes during a period for which we lack data via the procedure known as curve fitting. This technique is easily adapted for deriving projections, as will be shown in the following section.

Four models, the linear, geometric, exponential, and logistic, are commonly used, of which the first and third were introduced in Chapter 8 (also see Shryock and Siegel 1976: Chap. 13). Although there are infinitely many possible growth formulas, these four have been found most effective. One reason is that the curves associated with them are smooth. In the case of the first three, they assume that population growth occurs in a fairly even and regular manner rather than exhibiting dramatic peaks, valleys, reversals, and the like. In addition, records of population growth and growth rates taken from the Census Bureau and other sources verify that these give the most realistic picture of how populations actually change.

Arithmetic Growth

As we have seen, the linear model—also referred to as arithmetic growth—assumes that the rate of change between two dates will be constant throughout the interval. If the rate of growth at the beginning of a 10-year interval was 0.90 percent, then this model estimates growth as if the rate were 0.90 percent each and every year.

Arithmetic (linear) Growth: $P_t = P_0 + (P_0 \times GR \times t)$.

- P_t is the population size at a later date
- P_0 is the size at the earlier date
- GR is the growth rate
- t is the amount of time (number of years) between 0 and t

This is the same method that is used to determine simple interest on a savings account. A fixed rate is applied to the initial principal, say 10 percent per annum on $1,000, and that amount—$100, is added to the account each year. As shown in Figure 10.1, this equation traces a line to indicate the constant rate of change.

The population size of the United States was approximately 281,421,906 in midyear 2000, and the growth rate year was in fact 0.90 percent. Thus, if we wanted to estimate the population size in 2001, 2003, and 2005, when no Census counts were taken, we could use the linear model—provided that we were willing to make the constant-growth assumption. With t = 1, 3, and 5 for 2001, 2003, and 2005, respectively, the formulas would be:

$$P_{2001} = 281{,}421{,}906 + (281{,}421{,}906 \times 1 \times 0.009) = 283{,}954{,}703$$
$$P_{2003} = 281{,}421{,}906 + (281{,}421{,}906 \times 3 \times 0.009) = 289{,}020{,}297$$
$$P_{2005} = 281{,}421{,}906 + (281{,}421{,}906 \times 5 \times 0.009) = 294{,}085{,}891$$

To illustrate the use of the linear model between two dates for which we have data, we first need to rearrange the growth formula to solve for GR. This will then allow us to apply that rate (using the linear assumption) to any intervening year.

- First, we subtract P_0 from both sides, leaving: $P_t - P_0 = P_0 \times GR \times t$.
- Then, we divide both sides by P_0 and t, and switch sides of the equal sign.
- This gives us: $GR = (P_t - P_0)/(P_0 \times t)$.

Now, the enumerated U.S. population size in 2010 was just under 308,745,600, and we saw that the 2000 figure was 281,421,900. Thus the difference, or ($P_t - P_0$), is 27,323,700. With t = 10 years ($P_0 \times t$) = and GR, or $(P_t - P_0)/(P_0 \times t)$ = 27,323,700/2,814,219,000 = .0097 or 0.97 percent. Because this is slightly above the growth rate at the beginning of the interval, let us make some comparisons. We found that the 2005 estimate using the growth rate at one date (2000) was 294,085,891. But, if we apply the growth rate determined with two dates, then the 2005 estimate would be:

$$281{,}421{,}900 + (281{,}421{,}900 \times .0097 \times 5) = 295{,}070{,}862$$

Furthermore, since we know the 2010 enumeration total, we can compare that to the one-date estimate (with the rate of 0.90 percent) when t = 10 years. That is, $P_{2010} = P_{2000} + (P_{2000} \times .009 \times 10) = 281{,}421{,}900 + (281{,}421{,}900 \times .009 \times 10)\ 306{,}749{,}871$. Thus, we find a difference of about 1,995,700 persons between the estimate and the enumeration, with the estimate about 2 million too low because it did not account for a slight rise in the growth rate during the intercensal period.

Although the linear model did not prove to be perfect when applied to recent dates from the United States, it does appear to be fairly accurate. However, the assumption of a constant growth rate is generally inappropriate when we deal with more volatile

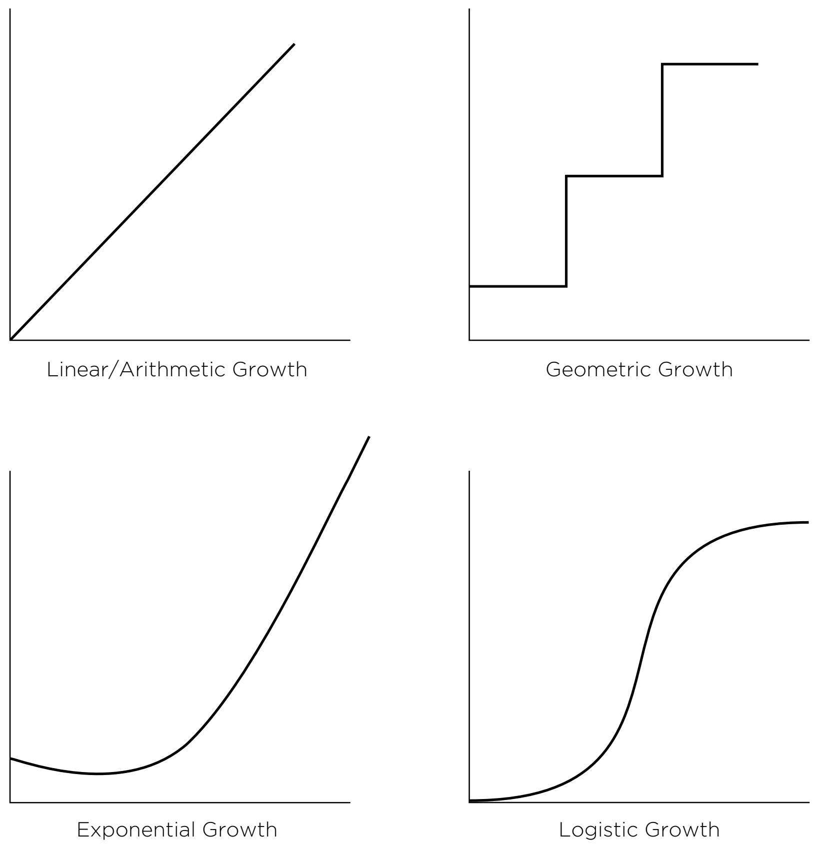

Figure 10.1. Four Simple Growth Models

populations and longer time periods. For example, in 1985 the population growth rate of Afghanistan was –2.8 percent per year, and its population size was 11,528,977. In 2010, the total population size was estimated at 28,397,812. During those intervening 25 years, the country, one of the poorest in the world, experienced:

- a civil war,
- a population explosion,
- a period of antinatalist policy, and
- an era of pronatalist Islamic fundamentalist family law, similar to Iran's under the Ayatollah Khomeini
- and a war with U.S. and Allied forces

Does the linear model explain its population growth? By the formula, the 2010 population size would be $P_{2010} = 11{,}528{,}977 + (11{,}528{,}977 \times -0.028 \times 25) = 3{,}458{,}693$, more than 25 million short. Obviously, Afghanistan's population did not grow arithmetically.

Geometric Growth

In Chapter 1, we noted that Thomas Malthus, who lived in an era during which his native England and other parts of Europe and America were experiencing civil war, revolution, and population explosions, argued that populations did not grow at the moderately slow pace depicted by the linear model. Rather, he believed that populations inevitably exhibited what he called "geometric" growth. That is, instead of expanding in an arithmetic series (with a constant rate of change) such as this: 1, 2, 3, 4, 5, and so on, he held that populations expanded in a geometric series (in which the rate of change itself increases), such as: 1, 2, 4, 8, 16, and so on. Although we now know that no single model or "law" applies to each and every population during any time period one chooses, it is true that **geometric growth** is often a more realistic assumption than the linear one. (It traces the step curve shown in Figure 10.1).

$$\text{Geometric Growth: } P_t = P_0 \times (1 + GR)^t$$

- P_t is the population size at a later date
- P_0 is the size at the earlier date
- GR is the growth rate
- t is the amount of time (number of years) between 0 and t

This model is applied to determine interest on savings compounded annually. A set interest rate, e.g., 10 percent, is applied to the initial principal, say $1,000, which at the end of the first year yields $100. For the second year, the rate is applied to the principal plus the interest, $1100, to yield $110 in interest, and so forth.

Let us illustrate, once again with recent U.S. Census data. Using the 2000 total of 281,421,900 and growth rate of 0.90 percent, we estimate the population sizes for 2001, 2003, 2005, and, for the sake of comparison, 2010.

P_{2001} =281,421,900 $(1 + .009)^1$ = 283,954,703
P_{2003} = 281,421,900 $(1 + .009)^3$ = 289,088,888
P_{2005} = 281,421,900 $(1 + .009)^5$ = 294,315,904
P_{2010} = 281,421,900 $(1 + .009)^{10}$ = 307,800,671

We can see that the geometric model yields estimates that are larger than those produced by the linear model (or larger in absolute value when growth is negative). The 2010 estimate of just over 307 million exceeds the linear estimate by about 1.05 million, and it is closer to the enumerated figure by about 0.9 million. This better, but still not perfect, estimate allows us to conclude that growth between 2000 and 2010 followed a course between linear and geometric, and closer to geometric.

Because geometric growth is more rapid than linear, it might be more helpful in explaining the situation in Afghanistan. Recall that the country's 1985 population size was about 11,528,977. Applying the geometric formula to estimate the 2010 total, we find that:

$$P_{2010} = 11{,}528{,}977 \times (1 - 2.8)^{25} = 944{,}867$$

which is still far below the estimate of more than 28 million. This leaves several possibilities, including that another model incorporating even a faster rate of increase would be more appropriate. Thus we are led to consider the third type of estimation curve, one that is believed to best fit populations undergoing explosive growth, the exponential model.

Exponential Growth

The assumption made by the geometric model, that increments ("interest payments") are added to the population base (the "principal") at the end of each year, is not realistic. In demographic applications, our attention is focused on vital events, births, deaths, and the comings and goings of migrants. These events occur continuously, hour by hour, day by day, and week by week. People don't wait until December 31st to bear children, or to pass away, or to change residence; yet, this is what geometric growth stipulates. When Malthus spoke of this kind of growth as being especially applicable to human populations, he almost certainly did not mean it literally. For the family of equations and curves based on the exponential model treats incremental (and decremental) factors as they actually occur. It is comparable to interest compounded daily, or even hourly, as close to momentarily as is practical. Most banks have adopted this method, and it is the kind of bank in which you want to keep your money.

Exponential growth traces the smooth, upward sloping curve in Figure 10.1. And its formula is

$$\text{Exponential Growth: } P_t = P_0 \times e^{GR \times t}$$

- P_t is the population size at a later date
- P_0 is the size at the earlier date
- GR is the growth rate
- t is the amount of time (number of years) between 0 and t
- e is the exponential constant = 2.71828

Just as we can determine the linear growth rate from its associated equation when both intercensal dates P_t and P_0 are known, exponential growth rates are determined by solving the preceding formula for GR. Using the algebraic concept of logarithms mentioned in Chapter 8.

$$GR = [\ln (P_t/P_0)]/t, \text{ where ln is the natural log (with the base e).}$$

Once again, we turn to the example of the U.S. population between 2000 and 2010. We have already seen that the geometric model assumes a rate faster than that actually achieved. Thus, we should expect that the estimates that anticipate exponential increase would be even higher and less accurate. With the 2000 population size and growth rate given, we solve for 2001, 2003, 2005, and 2010.

$$P_{2001} = 281{,}421{,}900 \times 2.71828^{(0.009 \times 1)} = 283{,}966{,}128$$
$$P_{2003} = 281{,}421{,}900 \times 2.71828^{(0.009 \times 3)} = 289{,}123{,}799$$
$$P_{2005} = 281{,}421{,}900 \times 2.71828^{(0.009 \times 5)} = 294{,}375{,}147$$
$$P_{2010} = 281{,}421{,}900 \times 2.71828^{(0.009 \times 10)} = 307{,}924{,}605$$

We see that these estimates are indeed the highest of the three sets, with the 2010 total close to the enumerated count.

Because all three sets of estimates based on the 2000 growth rate of 0.90 percent were inaccurate, it is most likely that actual growth in the 2000–2010 interval did not occur at that rate. When we assume exponential growth (a realistic model, as noted) we find that the associated growth rate is slightly below the 1980 figure. That is:

$$GR = [\ln (308{,}745{,}600/281{,}421{,}900)]/10 = 0.92 \text{ percent}$$

Recall that this is an annual average over 10 years.

When we fit Afghanistan's growth between 1985 and 2010 to the exponential model, we derive an estimate somewhat closer to the actual situation. For that country,

$$P_{2010} = P_{1985} \times 2.78128^{(-.028 \times 25)} = 5{,}764{,}488$$

Clearly, even the exponential estimate is low. In this case, it is reasonable to conclude that the growth rate at the beginning of the period of –2.80 percent did not continue at that level through to the end. In fact, we know that it fluctuated considerably with events. In 1985, it had gone to zero or even negative, at about –2.8 percent; by 1990, it had soared to 4.5 percent; in 1995 it was 6.5 percent; and in 2000 it was 3.00. Taking the exponential average for the period, we find that it is:

$$GR = [\ln (11{,}528{,}977/28{,}397{,}812)]/25 = 3.61 \text{ percent}$$

It is this rate which, when applied over the long range, that explains Afghanistan's uneven, but rapid, population explosion.

Logistic Growth

The logistic model assumes that neither absolute growth nor growth rates remain constant over the period under observation. Instead, it depicts a situation in which growth is rapid during the early part of the interval, then it slows until, during the last portion, it approaches or equals zero (or even goes negative). Its wide application in many fields in addition to demography has given the curve several names. One is the "S" curve, because of its shape. Another is the "saturation" curve because it is a model of how a permeable material, such as a sponge, fills up, taking on large amounts of liquid when it is dry but less as it fills and then none when it is saturated. A third is the "ogive," familiar to statisticians as the curve representing cumulative frequency. A fourth is "Pareto's curve" because it applies to a principle of economic saturation, Pareto's Law, developed by the early-twentieth-century Italian social scientist. It is also the mathematical model that best fits demographic transition, as Keyfitz (1968:76, 215) first suggested (also see Weinstein 1980:73).[4]

Table 10.2. Historical Population Data for the United States

Year	U.S. Population	Year	U.S. Population
1900	76,212,168	1960	179,323,798
1910	92,228,496	1970	203,302,031
1920	106,021,537	1980	226,542,199
1930	123,202,624	1990	248,709,873
1940	132,164,569	2000	281,421,906
1950	151,325,798	2010	308,745,538

Source: U.S. Census Bureau Fast Facts.

Box 10.1 Computations for Fitting Historical U.S. Population Sizes, 1900–2010, to the Logistic Curve

Notes: Computational Procedures

There are several computational procedures to choose from. The one used here is known as the method of selected points. Table 10.2 presents historical population data for 12 decades commencing from 1900. We selected three data points, 1900 (P_0), 1950 (P_5), and 2000 (P_{10}). The computation procedure involves estimating several coefficients which are used in the final projection formula.

1. Obtain the reciprocals for the selected years and multiply by 1 million.
   $(1/P_0)$ x 1,000,000 = 0.131213
   $(1/p_5)$ x 1,000,000 = 0.006608
   $(1/p_{10})$ x 1,000,000 = 0.003553
2. Calculate the difference between the first two reciprocals, and between the second and the third.
   $D_1 = (1/P_0) – (1/P_5) = 0.006513$
   $D_2 = (1/P_5) – (1/P_{10}) = 0.003055$
3. Obtain coefficient "a" = (1/r) x [ln(D_1)–ln(D_2)] , where "r" is 5 because the three points we chose are five units apart. a = (1/5) [–5.033909–5.790975] = 0.1514133.
4. Obtain the second coefficient, k, by first obtaining its reciprocal. $1/k = (1/P_0)–(D_1^2/(D_1–D_2))$ = 0.000861. Therefore, k = 1161.575091.
5. Obtain the third coefficient, b. b = [(k/P_0)–1] = 14.241333.
6. We will project the U.S. population to year 2015 first.
   $P_{(2015)} = (k/[(1+b) \times (e^{-aT})]$ where T is (2015–1900)/5
   $P_{(2015)}$ = 1161.57509/3.314488 = 350,453,803 July 2015

The U.S. Population Clock estimates the population midday March 26, 2015 as 320,577,213.
P (2020) = 388,579,308.

The curve representing **logistic growth** is shown in the lower right panel of Figure 10.1, and its formula is:

$$\text{Logistic Growth: } P_t = k/[(1 + b) \times (e^{-aT})]$$

- P_1 is the population size at a selected date (for which data are unavailable)
- k is an estimate of the largest population size attainable over the observation period, based on but usually larger than P_0
- b is another estimated constant that represents the length of time between P_0 and the point at which growth begins to slow
- e is the exponential constant (indicating that this curve is related to the one for exponential growth

- a is an estimated average rate of growth for the entire period
- T is the number of five year segments (our population data points are five years apart) in the duration to the projected date

Table 10.2 shows the results of applying this formula to the U.S. population, with the details of the calculations provided in the notes.

In Chapter 8, we saw how this type of change actually occurred in most of the populations in Europe between about 1700 and 1900. Just prior to the Industrial Revolution, following millennia of slow, uneven, "Malthusian" growth, general mortality rates and IMRs began to decline in England, France, and eventually other countries. As the effects of industrialization spread, this process accelerated, resulting in the highest rates of natural increase ever experienced in any human populations to that time. Then, around 1850, with the increase in the size and influence of the urban middle classes and the growing popularity of the family-planning movement, birth rates began to decline. They have continued to do so throughout Europe, the United States, and other parts of the world. With these declines came progressively slower population growth until, today, many countries are at or below ZPG.

One country that now appears to be in the midst of its transition is Brazil. With a total population size of 169 million in 1998[5] and a rate of natural increase of 1.5 percent, it is growing more rapidly than the United States and the European countries that have essentially completed their transitions. But it is growing more slowly than some of its neighbors such as Bolivia, Colombia, Paraguay, Ecuador, and Peru—and much more slowly than many Asian and African nations. Of even greater significance, Brazil's population growth rates have declined substantially over the past few decades when, as recently as 1960, the rate was nearly 3 percent.

Components Methods

Methods based on the fundamental equation of demography apply the fact that the size of a population at a date for which data are unavailable can be estimated. This requires information about (1) known population size at an earlier or a later date—or both—and (2) the volume of births, deaths, in-migrations, and out-migrations that occur between dates. The simplest of these, the **cohort survival method**, assumes a closed population in which net migration is set at zero. Under these conditions,

$$P_t = P_0 + B - D \text{ or } P_0 = P_t - B + D$$

where B is the number of births that occur between 0 and t, and D is the number of deaths in the interval. Note that we can either work forward from an earlier date, adding births and subtracting deaths, or from a later to earlier date by subtracting births and adding deaths. The assumption of a closed population does apply to the world, which experiences neither in- nor out-migration (thus far). It also characterizes populations in which there is negligible or no movement in or out, and it fairly closely approximates those in which in-migration exactly equals out-migration.[6] In general, when applied to today's dynamic national aggregate, state, provincial, and urban populations, this method is prone to inaccuracy, for which adjustments must be made.

Cohort Survival

This issue aside, the cohort survival method is more precise than the simple growth models because in the place of crude rates it employs age-specific measures, in a

Table 10.3. Components Method for Estimating Population Size

	Initial Population (P_0)		Survival Rates			Estimated Population (P_t)	
Age	Males	Females	Males	Females	ASFR	Males	Females
0–9	20709.0	19841.0	0.995	0.996		23045.00	22141.00
10–19	21884.0	20834.0	0.975	0.980	0.04	20605.45	19761.63
20–29	21650.0	21038.0	0.965	0.970	1.60	21336.89	20417.32
30–39	20039.0	20104.0	0.950	0.960	0.51	20892.25	20406.86
40–49	21603.0	21997.0	0.930	0.950	0.02	19037.05	19299.84
50–59	20457.0	21506.0	0.920	0.940		20090.79	20897.14
60–69	18175.0	20357.0	0.900	0.930		18820.44	20215.64
70+	7267.0	11288.0	0.890	0.910		22824.00	29204.0

Note: Survival refers to forward survival. Thus, a woman's probability of surviving from ages 20–29 to 30–39 is .0.97, etc. This hypothetical model assumes higher survival rates for females at all ages and no fertility before age 15 or after age 49.

Source: P_0 based on data from Howden and Meyer (2010: Table 2).

manner similar to life table analysis. The three basic variables are: (1) the age-sex structure of the population: the size of each cohort of males and females taken separately, (2) the age and sex-specific survival rates, derived from the life table, and (3) age-specific fertility rates (ASFRs). The data can be unabridged, using one-year cohorts, 0–1, 1–2, and so forth; or, more commonly, abridged, using five- or 10-year cohorts (except for the 0–1 group). In the latter case, the highest age category is open and includes all individuals ages 65 and above, 70 and above, or 85 and above. Survival is measured by the forward rates (S_x), which indicate the probability of surviving from one age to another, for example, from 35 to 40.

The estimation procedure is illustrated in Table 10.3, which shows an initial population, P_0—that of the United States as enumerated in 2010, divided into cohorts for males and females separately. The total size is 308,749,000 with 151,784,000 males and 156,965,000 females. **Cohort survival** is employed to estimate the size of each cohort and of the population 10 years later, P_t (because we are using 10-year cohorts). Once this is accomplished, the exponential growth model is used to estimate the population size in various intervening years. The survival and ASFRs are realistic but hypothetical.

The procedure occurs in three steps. First, we obtain the estimated number of children in the 0–9 cohort in P_t. This is done by multiplying the age specific birth rates by the female population of each cohort with a fertility rate above 0 (column 4) x (column 7), and summing the products. This sum is then multiplied by 0.51 to obtain the male births and by 0.49 to obtain the number of female births. (This assumes that there are 510 males in every 1,000 births.)

The second step is to derive the size of each of the remaining cohorts in P_t, from ages 10–19 to 60–69 (the size of the last cohort with an open age interval is calculated separately). This is accomplished by applying the survival rates to the respective cohorts, multiplying column 3 by column 5 to obtain the estimated age-specific male population, and column 4 by column 6 for the female population. For example, to estimate the number of 20–29-year-old males in P_t, the 10–19 cohort in P_0 is diminished by applying its 10-year survival rate. That is, 21,084,000 × .975 = 206,050,000.

The third step is to obtain the estimated size of the last age cohort. In P_t, this group will contain the base population in the age group 60–69 surviving to the next age category plus those in the 70+ category surviving the 10 years between time 0 and time t. For example, the projected female population age 70+ is obtained as follows: (20,357,000 × 0.930) + (11,288,000 × 0.910) = 29,204,000.

This procedure yielded a total size for P_t of 338,994,000, of whom 166,651,000 are male and 172,343,000 are female. Now, with two population sizes at two different dates, one actual and the other estimated, we can derive estimates for intervening dates. The exponential growth model indicates that the average annual growth rate for the interval, based on our assumptions, is:

$$GR = [\ln (338{,}994{,}000/308{,}345{,}000)]/10 = .00947 \text{ or } 0.947 \text{ percent}$$

Applying this rate in the formula:

$$P_t = P_0 \times e^{GR \times t}$$

where t = 2011, 2012, 2013, and 2014, we derive these estimates:

P_{2011} = 311,278,000
P_{2012} = 317,504,000
P_{2013} = 323,730,000
P_{2014} = 329,955,000

Because this method uses age- and sex-specific data, estimates can be derived not only for total population sizes but also for the sizes of cohorts, male, female, and combined, at dates between the base and terminal years. Two cohorts of special interest in planning and administrative contexts, one a school-age group (10–19) and the other seniors (70+), were selected and their sizes estimated for the same four dates, also assuming exponential growth. These are:

10–19 cohort: P_{2010} = 42,717,000	70 + cohort: P_{2010} = 14,555,000
P_{2011} = 43,123,000	P_{2011} = 18,731,000
P_{2012} = 43,985,000	P_{2012} = 19,106,000
P_{2013} = 44,848,000	P_{2013} = 19,480,000
P_{2014} = 48,710,000	P_{2014} = 19,885,000

When we observe the estimates for the entire aggregate and the cohort sizes together, we see clearly how a population simultaneously grows and gets older, as is the case in the United States and other industrialized nations.

Components Methods that Include Migration

The more complex components methods take migration into consideration. In preparing its estimates, the U.S. Census Bureau uses this approach, choosing among several models that differ from each other with respect to the techniques used to estimate the net migration component.[7] These include (1) the components II method, (2) the administrative records method, and (3) the ratio correlation method. This section summarizes the first two of these approaches.

1. The components II method is widely used (Ferlay et al. 2013; CPR 1976; Raymondo 1992) and can be applied to national aggregates, regions, states, and other local areas. Like the cohort survival method, it begins by establishing the size and age-sex distribution of the central-year population from the most recent census data available. In this case, however, only the civilian population under age 65 is considered; that is, special groups such as members of the military, long-term hospital residents, and college dormitory residents are subtracted from the base total. Natural increase is derived by subtracting the total number of deaths from the total number of births during the period, using data obtained from vital-statistics registers.

2. The administrative records method is used by the Census Bureau to estimate the sizes of state, county, and subcounty (e.g., urban) populations. Like the components II method, it employs the demographic accounting formula. Thus it requires information about (1) population size for the estimate area (state, county, and so on) at a base date, and (2) the number of births, deaths, in-migrations, and out-migrations in the interval between the base date and the date for which an estimate is to be determined. These vital statistics are taken from the records of non-Census administrative organizations, and it is this practice that gives the method its name.

Data on deaths and births are obtained from the NCHS and the State health departments via the Census Bureau's Federal-State Cooperative Program for Population Estimates (FSCPE). The number of internal migrants is determined using Federal income tax returns provided by the Internal Revenue Service. Tax returns include the social security number and the address of the person filing, which can be compared at the base and estimate years. Those who file from two different addresses during the estimation period are classed as movers or as migrants, depending on whether the change of address is within a county (mover) or between counties (migrant), or

Table 10.4. Estimates Based on Components Method, 2013–2014 for the United States, Regions, and Selected States

	Population 2013	Births	Deaths	NIM	NDM	Change	Population 2014
United States	316,128,839	3,957,577	2,593,996	995,944	0	2,359,525	318,488,364
Northeast	55,943,073	637,853	478,007	262,204	-286,696	124,113	56,067,186
Midwest	67,547,890	829,620	586,099	127,607	-182,057	177,237	67,725,127
South	118,383,453	1,511,280	1,007,640	358,956	365,289	1,249,132	119,632,585
West	74,254,423	978,824	522,250	247,177	103,464	809,043	75,063,466
Arizona	6,626,624	86,868	51,748	14,234	41,975	96,487	6,723,111
California	38,332,521	505,903	255,787	161,318	-32,090	371,107	38,703,628
Florida	5,268,367	214,567	187,102	112,306	138,546	292,986	5,561,353
Michigan	19,552,860	112,748	90,366	20,094	-28,679	11,684	19,564,544
Nevada	9,895,622	35,153	21,702	8,456	23,623	47,605	9,943,227
N. Dakota	2,790,136	10,780	5,879	1,290	8,974	15,625	2,805,761
Pennsylvania	723,393	142,032	128,600	29,060	-31,448	5,913	729,306
Rhode Island	12,773,801	10,918	9,543	4,290	-3,387	1,819	1,277,562
West Virginia	1,051,511	20,466	21,735	1,164	-2,749	-3,269	1,048,242

Notes: NIM is net international migration, NDM is net domestic migration.

Source: U.S. Bureau of the Census: State Totals Vintage 2013 Estimates of the Components of Resident Population Change for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2013.