Chapter 3 Demographic Analysis

3.1 The Need for Demographic Analysis

Planning for the future requires, to some extent, making projections based on past observations The U.S. Census Bureau provides, as a routine procedure, national and state-level population projections.[①] State governments, often in cooperation with an external agency such as a university, do more geographically focused population analyses and projections. For example, the Urban Studies Institute at the University of Louisville, part of the Kentucky State Data Center (KSDC), is responsible for the periodical projection of future population trends at the state and county-level and for selected cities in Kentucky.[②] Dealing with the uncertainty of future estimated births, deaths, and migration patterns, the institute offers three simultaneous population projections at low, middle and high growth rates. Additionally, the institute makes a variety of past and present population estimates available online. For more geographically detailed population projections and estimates, local government agencies, such as city planning departments or county planning commissions, engage in all sorts of methods to evaluate past and present demographic trends.[③]

Generally, population projections are the base for many planning activities, such as producing land use and transportation plans, determining the direction of future economic development and providing guidance for housing, school, and shopping center developments. Population projections often become the center-piece of comprehensive plans and the future vision of localities. The importance of population estimates and projections in planning becomes apparent by looking at some selected, local planning issues:

Land use planning: General land use and specific development policies need to regularly address increasing population size. A town's conceptual image and physical appearance depends largely on future land use planning. Expected population growth patterns drive much of the decision-making processes such as designating more residential areas; finding the right combination of residential,

---

① Schedule of population and household projection releases by the U.S. Census Bureau: http://www.census.gov/population/www/projections/projsched.html. State Population projections by the U.S. Census Bureau: http://www.census.gov/population/www/projections/stproj.html.

② Urban Studies Institute at the University of Louisville: http://ksdc.louisville.edu/Projections2003.htm.

③ Boone County, Kentucky, official website: http://www.boonecountyky.org/.

commercial, office, and industrial uses in mixed-use areas; and allocating parks and open spaces.

Transportation planning: Growing cities and metropolitan regions face the challenge of coping with increases in transportation demand. More automobiles on the streets and highways and higher demands for public transportation systems have their origins in growing populations.

Economic development: A growing economy creating sufficient employment opportunities, which in return allows a sustained increase in people's standard of living is central to economic development planning. A region's population and its growth trend is thus of major importance. For instance, a growing number of people will foster local retail sales, will be the basis of a qualified labor pool for expanding industries, and will be the basis of various tax incomes for state/local governments.①

Environmental planning: Planners constantly face the challenge of preserving nature and wildlife habitat while providing high quality spaces to meet the demand for human activities. Population analysis provides the base for searching the balance between human and nature.

Housing: Booming regions have tremendous demand for housing. Knowing the projected population increase for a specific time period will give some guidance to those who must accommodate this demand. Identifying demographic characteristics, such as persons per household, will add valuable information on the total future need of housing units.

Public services and facilities: Imagine that public services cannot keep pace with population increase. The direct result would be garbage-filled streets and bottlenecks in the provision of water and electricity. Planning ahead for anticipated population growth is essential in public services and facility plans.

Sustainable development: Sustainability can be defined as finding a level of economic development that does not compromise the economic vitality of future generations or the integrity of the natural environment. Population growth that, for instance, considers resource requirements environmental constraints would be a first step towards building a healthy and sustainable urban and regional environment. Translating the idea of "*Constrained Economic Growth*"②, into holistic urban and regional planning would lead to sustainable community development. In the long run, intergenerational equity would be reached, in

---

① Following the principles of the economic base theory, increasing export demand for regionally produced goods and services may also contributes to a large extend to a region's economic prosperity.

② Batie Sandra, 1989: 1,084 – 1,085. "Sustainable Development: Challenges to the Profession of Agricultural Economics." *American Journal of Agricultural Economics*, December: 1,083 – 1,101. We recognize that advocates of the "Resource Maintenance Definition" of sustainable development among others argue for the separate maintenance of human and natural capital, since they are complements rather than substitutes (Daly and Cobb, 1989:72). The degree to which this separation is handled leads to the distinction of "weak" and "strong" sustainability.

which the demand of growing population is addressed with the most appropriate measures in transportation, land use, and economic plans that always consider environmental quality.

All these planning examples demonstrate the importance of understanding past and present demographic population characteristics, such as gender and age distributions and expectations of future population development. They emphasize how the analysis of past and present population statistics and future population developments can play a key role in a variety of planning and decision processes. For service deliveries by local governments, in transportation, land use or environmental planning, underestimating future populations can lead to shortages and a reduction of the quality of life. Overestimating future populations, on the other hand, may result in wasting local resources through costly oversupply of services.

Before we actually start analyzing past and present population characteristics, we first need to discuss the terms and definitions used by demographers and the components of changes in population trends.

3.1.1 Typology of Projection Methods

There is a wide range of population projection methods in the literature and there are several ways of classifying population projection methods. As a first distinction, we can clearly make a difference between **subjective** and **objective projections** (Armstrong, 1985). Subjective projections can be simply described as "wild guesses" and as such abstain from a rigorous systematic and methodological approach. They depend largely on feelings and intuitions and, at their best, can only reflect impressions on future population tendencies. Objective projections follow the quantitative approach of collecting data and applying a quantitative method to obtain a projected result.

Given the importance of population projections for the planning community, you can easily imagine that you would not risk your planning career by depending entirely on subjective projections. However, we must recognize that objective projections also rely, to some extent, on subjective elements. This includes the choice of the "correct" projection method and/or as mentioned before, the selection of the right parameters used in projections.

Some projection methods depend solely on historical trends (e.g., trend extrapolation models) while others account for various interrelationships of population statistics with non-demographic variables, such as regional employment, amenities and wage levels (e.g., structural population models). Figure 3.1 identifies three main population projection methods: trend extrapolation, cohort-component, and structural.

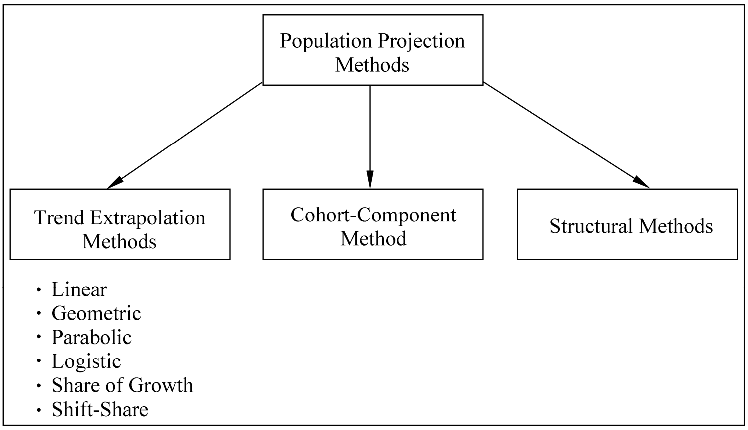

Figure 3.1 Population projection methods

Trend extrapolation methods observe historical trends and project them into the future. These methods are often used for small areas where disaggregated population statistics are not always available. They are powerful straightforward tools for projecting populations because they rely on a single, highly aggregated, data series. For instance, we can extrapolate the observed population trend for the past twenty years for Cincinnati into the near future. The different individual methods listed under this group (e.g., linear, geometric, etc.) refer simply to different mathematical approaches of finding the best fit for the observed data. An important thing to remember is that trend extrapolation does not account for any causes of these past observed trends. This is where the **cohort-component** [①] methods come into play. The most common version of the cohort-component methods uses sex-age-specific population cohorts and adjusts them for the three factors of population growth: births, deaths, and migration. Subdividing the sex-age-specific cohorts further by race/ethnicity increases the level of detail, but also increases the data requirements. The level of detail and the fact that it accounts for the components of population changes make this group the most frequently used population projection method. Accounting further for non-demographic factors leads us to the **structural models**. Beyond the scope of this textbook and often very complex in nature, models falling into this group explain population growth (dependent variable) through a variety of non-demographic (independent) variables such as employment, wage levels, and local amenities as well as land use and transportation models.[②]

Now that we have briefly described the three major categories of population

① According to relevant literature, the origins of the cohort-component method go back to Refs. (Carnan, 1985; Bowley, 1924; whelpton, 1928).

② A detailed discussion of structural models is offered by Smith et al. (2001).

projection methods, we will take a closer look at factors that affect the choice of methods. Each method represents a unique mix of characteristics, assumptions, and requirements. So how can you be sure to pick the most appropriate projection method? Surely, understanding the unique characteristics, assumptions, and requirements is an essential step towards making an educated choice, but there are many more factors that must be considered.

Subjective impressions: Although you have decided to go with an objective population projection, you may have a tendency of using one method over others. This might be because a method appears more elegant, more reliable, or you simply prefer to copy a “similar” study from a neighboring/close-by county for which you have a detailed description of the method.

Time constraint: Usually, people expect you to get the job done within a certain time. This is no exception when it comes down to doing a population projection. Your judgment on how long it will take to collect the data, do the analysis, and write a report will certainly influence your decision on what method to choose.

Technical skill level: People tend to avoid methods with which they feel uncomfortable with. Lack of adequate training, for example, can be one reason to choose a more straightforward approach over a more complex one.

Data availability: As a general rule, data are more widely and easily available for larger geographic regions. For the United States, the Census Bureau offers, for example, population estimates by age, sex, race, and Hispanic origin at the national, state, county, and sub-county level, such as census tract and block group.[①] State governments usually maintain population statistics on a regular basis, often associated with state universities. Generally, trend extrapolation methods usually have lower data requirements when projecting population totals. The cohort-component method on the other hand has higher data requirements by using sex-age-specific population cohorts.

Detail of analysis: Are you interested in total population changes or do you need to analyze the underlying causes for these population changes, such as migration, births, and deaths? The level of detail in your analysis can play a major role when choosing among simpler trend extrapolation techniques or the more data-intensive cohort-component model.

Purpose of the population projection: It makes a big difference if your supervisor asks in an informal way for a rough figure as a base for follow-up analyses or if your analysis will be posted on the county’s planning commission website as part of the comprehensive plan.

Strengths and weaknesses: Every method has strengths (e.g., low data requirement), which may at the same time lead to weaknesses (e.g., low level of detail). Balancing the pros and cons might provide further guidance of what method might be most appropriate.

---

① Source: U.S. Census Bureau, Population Division: http://eire.census.gov/popest/estimates.php.

3.2 Demographic Analysis—Fundamental Concepts

For planners and demographers alike, population analyses do not begin with immediately applying sophisticated methods in population projections. Rather, most demographic analyses start with fundamental concepts, including:

(1) describing populations by their actual size,

(2) determining population distribution across predefined areas,

(3) creating sex, race, and age composition profiles of populations of interest, and

(4) calculating observable percent changes of selected population characteristics.

The point here is to get a thorough understanding of the population of interest by studying characteristics for periods where data are available. For planning purposes, these first demographic analyses can already give planners valuable and necessary information. Is the population of an area declining or increasing? By what rate is the area declining or increasing? With such information, school district superintendents could make some educated guesses about expected enrollment if they know the age composition of the area's population. For the provision of public services, such as police and fire protection or the local library, local governments use population statistics to avoid costly over or under provisions of needed services. This list of examples on how population statistics influence the planning decision process could be extended.

Let us focus on the first of the fundamental concepts of demographic analysis, the **population size**. Using 1990 and 2000 U.S. Census Bureau population statistics for Boone County we see immediately that the county is growing fast. Boone County had a population of 57,589 in 1990 and 85,991 in 2000.

While the concept of population size is straightforward, it is an important fact that, in general, people are counted according to their permanent place of residence. For example, someone living in a neighboring county and commuting daily to Boone County for work is not considered a resident of Boone County. As a result this person would, of course, not show up in Boone County's population size in Table 3.1. The so called "de jure" approach counts people only at their permanent place of residence.

Table 3.1 Boone County population size, 1990 and 2000

Boone County, Kentucky	1990	2000	Absolute Change	Percent Change
Male	28,111	42,499	14,388	51.2
Female	29,478	43,492	14,014	47.5
Total	57,589	85,991	28,402	49.3

Source: U.S. Census Bureau, Data Set: 1990 and 2000 Summary Tape File 1 (STF 1)

The next demographic concept deals with **population changes**. Generally, change can be expressed as: absolute change, percent change, average annual absolute change, or average annual percent change. In Chapter 2, we used the population totals for Boone County for 1990 and 2000 from Table 3.1 above (e.g., 57,589 and 85,991 respectively) to calculated the four different measures of changes:

Absolute change: subtract the 1990 population from the 2000 population:

$$85{,}991 - 57{,}589 = 28{,}402$$

Percent change: divide the absolute population change by the 1990 population to get percentages:

$$28{,}402 / 57{,}589 = 49.3\%$$

Average annual absolute change (AAAC): divide the absolute population change by the number of years between 1990 and 2000; here we have exactly 10 years:

$$\text{AAAC} = 28{,}402 / 10 = 2{,}840$$

Average annual percent change (AAPC): apply the geometric growth formula, Eq. (3.1) and solve for the growth rate, Eq. (3.2): [①]

$$\text{Pop}_{2000} = \text{Pop}_{1990}(1 + \text{AAPC})^{\text{Years}} \tag{3.1}$$

$$\begin{aligned} &\text{AAPC}=[\text{Pop}_{2000} / \text{Pop}_{1990}]^{1/\text{Years}} - 1 \\ &\text{AAPC}=[85{,}991/57{,}589]^{1/10} - 1 = 1.0409 - 1 = 4.09\% \end{aligned} \tag{3.2}$$

The next concept is **spatial distribution of population**, the spatial pattern of human settlements. We all know that human settlements are not evenly distributed. For example, California (35,484,453) is one of the most populated states, while North Dakota (633,837), South Dakota (764,309), Montana (917,621), and Wyoming (501,242) belong to the least populated states in the United States.[②] Another way of expressing uneven spatial distribution of population is in the form of population densities, usually defined as persons per unit area. These examples show the population distribution across political areas, namely states. Other political entities may include counties, cities, townships and school districts. Common non-political, geographic entities include the various statistical entities used by the U.S. Census Bureau. The smallest geographic unit for which

① Most of you will recognize that the process of computing annual growth rates is identical to the more familiar process of compounding in finance or accounting. Here, instead of population data, we would use future and present values to represent the compound rate.

② The numbers in parenthesis are the U.S. Census Bureau population estimates for 2003.

the Census Bureau tabulates 100-percent data is the census block. Several blocks clustered together form a block group, and the next higher level is census tract. Effective June 6, 2003, the Census Bureau began using metropolitan and micropolitan statistical areas within a "Core Based Statistical Area" (CBSA) classification when referring to larger urban agglomerations. Metropolitan statistical areas must have at least one urbanized area with a population of 50,000 or more. Micropolitan statistical areas must have at least one urban cluster with a population of at least 10,000 but less than 50,000.[①,②]

The map in Fig. 3.2 shows how the Year 2000 population is distributed in Boone County. It illustrates that the population is not evenly distributed

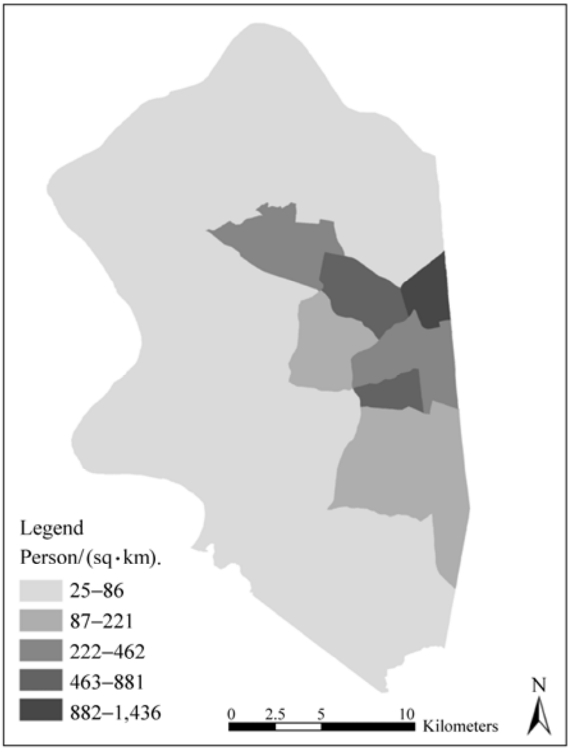

Figure 3.2 Boone County population distribution

① http://www.census.gov/population/www/estimates/00-32997.pdf.

② The six New England states use the same criteria for metropolitan and micropolitan statistical area definitions. A New England city and town area (NECTA) must have an urban core with a population of at least 2.5 million. Further subdivision of NECTAs is possible and referred to as New England city and town area divisions. The definition of boundaries is important in defining the area of interest for population analyses. Source: http://www.census.gov/population/www/estimates/metrodef.html.

throughout the county. A large share of the county's population (e.g., 23,551) lives in the east.

The last basic demographic concept refers to **population composition**. Commonly used population composition includes age and sex. Table 3.2 and Table 3.3 show the population composition for Boone County by age, sex, and race for the years 2000 and 1990.

Table 3.2 Boone County population by age, sex, and race, 2000

Boone County, Kentucky	White	Black or African-American	American Indian and Alaska Native	Asian including Pacific Islanders	Some other race, two or more races	Total
Total	81,822	1,306	200	1,137	1,526	85,991
Male:	40,347	716	88	544	804	42,499
0 to 4 years	3,240	60	8	57	137	3,502
5 to 9 years	3,461	67	7	40	110	3,685
10 to 14 years	3,301	61	8	40	67	3,477
15 to 19 years	3,023	34	7	24	54	3,142
20 to 24 years	2,421	67	4	23	76	2,591
25 to 29 years	2,787	82	7	42	85	3,003
30 to 34 years	3,350	73	8	84	69	3,584
35 to 39 years	3,618	69	8	82	75	3,852
40 to 44 years	3,578	77	10	47	37	3,749
45 to 49 years	3,020	61	10	39	23	3,153
50 to 54 years	2,520	27	6	28	29	2,610
55 to 59 years	1,887	22	2	15	14	1,940
60 to 64 years	1,344	3	2	8	8	1,365
65 to 69 years	1,020	9	1	9	8	1,047
70 to 74 years	829	1	0	5	4	839
75 to 79 years	553	1	0	1	5	560
80 to 84 years	243	0	0	0	3	246
85 years and over	152	2	0	0	0	154
Female:	41,475	590	112	593	722	43,492
0 to 4 years	3,065	71	11	61	139	3,347
5 to 9 years	3,242	59	7	46	104	3,458
10 to 14 years	3,151	44	12	32	70	3,309
15 to 19 years	2,811	36	8	22	63	2,940
20 to 24 years	2,361	45	9	23	52	2,490
25 to 29 years	2,921	47	7	56	69	3,100
30 to 34 years	3,411	53	10	86	61	3,621
35 to 39 years	3,804	55	11	84	49	4,003
40 to 44 years	3,699	66	9	70	35	3,879

Continued

Boone County, Kentucky	White	Black or African-American	American Indian and Alaska Native	Asian including Pacific Islanders	Some other race, two or more races	Total
45 to 49 years	3,087	41	9	34	17	3,188
50 to 54 years	2,646	26	6	30	22	2,730
55 to 59 years	1,862	15	5	13	18	1,913
60 to 64 years	1,388	9	4	15	3	1,419
65 to 69 years	1,156	7	3	9	7	1,182
70 to 74 years	1,092	3	1	8	6	1,110
75 to 79 years	773	5	0	2	4	784
80 to 84 years	522	3	0	1	1	527
85 years and over	484	5	0	1	2	492

Source: U.S. Census Bureau, Census 2000

We see immediately that we have one sub-table each for the male and the female population. This **sex (i.e., gender)** composition in Table 3.3 shows that there were slightly more females (e.g., 43,492) than males (e.g., 42,499) in Boone County in 2000.

Table 3.3 Boone County population by age, sex, and race, 1990

Boone County, Kentucky-Males	White	Black or African-American	American Indian and Alaska Native	Asian including Pacific Islanders	Some other race, two or more races	Total
Total	56,716	361	88	355	69	57,589
Male:	27,651	197	49	179	35	28,111
0 to 4 years	2,330	17	2	33	4	2,386
5 to 9 years	2,547	14	5	16	6	2,588
10 to 14 years	2,420	17	4	10	2	2,453
15 to 19 years	2,079	15	3	11	2	2,110
20 to 24 years	1,834	20	5	8	1	1,868
25 to 29 years	2,261	22	7	16	3	2,309
30 to 34 years	2,672	23	0	25	6	2,726
35 to 39 years	2,480	22	4	22	3	2,531
40 to 44 years	2,176	7	7	16	4	2,210
45 to 49 years	1,687	16	6	7	2	1,718
50 to 54 years	1,290	2	1	6	1	1,300
55 to 59 years	1,089	8	2	3	1	1,103
60 to 64 years	964	3	1	2	0	970
65 to 69 years	773	3	1	2	0	779
70 to 74 years	466	4	1	1	0	472

Continued

Boone County, Kentucky-Males	White	Black or African-American	American Indian and Alaska Native	Asian including Pacific Islanders	Some other race, two or more races	Total
75 to 79 years	282	1	0	0	0	283
80 to 84 years	177	3	0	0	0	180
85 years and over	124	0	0	1	0	125
Female:	29,065	164	39	176	34	29,478
0 to 4 years	2,249	12	2	23	3	2,289
5 to 9 years	2,339	13	1	16	5	2,374
10 to 14 years	2,282	14	6	6	6	2,314
15 to 19 years	2,003	9	3	7	5	2,027
20 to 24 years	1,915	11	2	12	1	1,941
25 to 29 years	2,480	16	2	20	3	2,521
30 to 34 years	2,935	16	1	39	4	2,995
35 to 39 years	2,527	20	6	10	3	2,566
40 to 44 years	2,256	13	2	19	3	2,293
45 to 49 years	1,665	10	3	6	1	1,685
50 to 54 years	1,277	5	4	6	0	1,292
55 to 59 years	1,152	8	3	4	0	1,167
60 and 61 years	1,093	2	3	5	0	1,103
65 to 69 years	912	5	1	1	0	919
70 to 74 years	683	3	0	1	0	687
75 to 79 years	541	4	0	1	0	546
80 to 84 years	432	2	0	0	0	434
85 years and over	324	1	0	0	0	325

Source: U.S. Census Bureau, Census 1990

Of major importance for planning purposes is the **age** composition of a population. As we mentioned already, demand for public services such as education, depends largely on the age structure of an area's population. Children and teenagers need to go to school and go to the playground after school. Young professional families have children and buy their first home. As people age, preferences and demands for public services change. Older people have usually higher demand for health care and nursing homes. We can easily see that the age structure of a population reveals important information on needs and demands of the population for planning purposes.

One term people often use to describe age groups is **age cohorts**. Usually, many demographic studies divide the population into five- or ten-year age cohorts. It reduces the number of total cohorts significantly compared to using one-year age cohorts. But it also reduces the level of detail by aggregating single years into multi-year age cohorts. In the example, the population of Boone County is divided into five-year age cohorts with the exception that the last age cohort includes all persons 85 years of age and over. Reading down the column labeled "total" for the female sub-table of Table 3.2, we can identify the following:

— the youngest age cohort 0 – 4 years has 3,347 members,
— the age cohort 35 – 39 years is the largest one with 4,003 females and
— the oldest age cohort of 85 years and over has 492 females.

Depending on what information you need, a table like this can provide the number of females of school age, the number of women of working age and the number of potential retirees living in the county.

Together, the **age-sex** composition of a population is often graphically represented in what is called a **population pyramid**. It is a double histogram of the sex-age structure where females are on the left side of the vertical zero line and males are on the right side. Each horizontal bar represents one age cohort, with the youngest age cohort at the bottom and the oldest age cohort at the top. The length of each bar is directly related to the number of persons it represents.

From Fig. 3.3, we see that there were more males than females in the youngest age cohort in Boone County, which is consistent with the worldwide observable phenomena that more males are born than females. The exact male/ female birth ratio is 1.05 stating that for every 100 female babies, there are 105 male babies born. The largest age group for both sexes in Boone County is that of age 35 – 39; thereafter the population cohorts decrease. Beginning with the age cohort 25 – 29, the number of females outweighs the number of males for all subsequent age cohorts, reflecting lower mortality rates for females at all ages. Also beginning

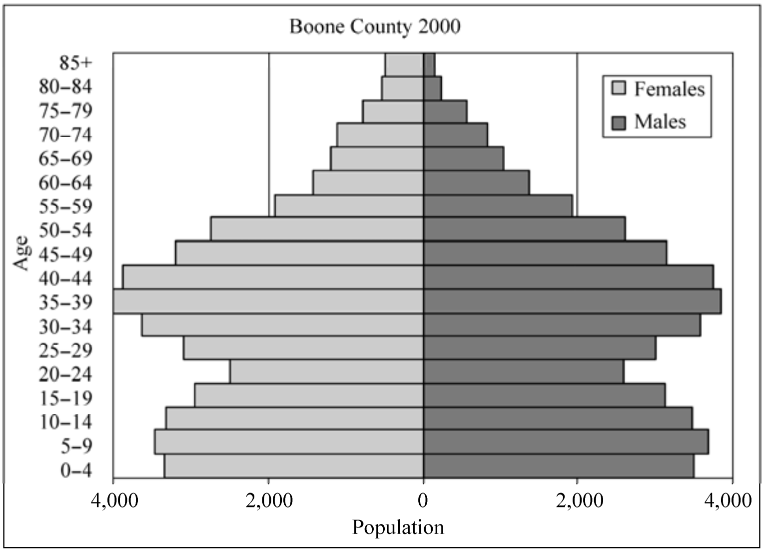

Figure 3.3 Population pyramid, Boone County, 2000

with the age cohort 60 – 64, the number of persons per age cohort declines drastically, partly due to exponentially increasing mortality rates in these upper age cohorts. Another factor that contributes to this is maybe the fact that many older people move to retirement homes that are away from Boone County.

For interpreting county-level population pyramids one must be aware of two forces that simultaneously shape the form of pyramids. One is the **fertility rate**, which plays an important role in projecting population. Although national fertility rates are easily available in the United States, fertility rates also show region-specific variances.

A second, and for smaller areas, more important force is **in- and out-migration**. The irregularity in the population pyramid for Boone County beginning with the age cohorts 15 – 19 can mainly be explained by age cohort specific migration rates. Later in this chapter you will see that this irregularity in the population age structure overlaps flawlessly with the county's age-specific migration rates. In other words, the observable decline, which begins with the age cohorts 15 – 19, can be attributed largely to migration.

This section closes with the realization that describing one area's population according to its size, distribution, composition, and change will reveal a lot of information. These four basic population characteristics should give you enough guidance to avoid unnecessary and wrong conclusions in population projections. We recommend studying all possible aspects of your target population before actually doing the projections. The more you know about the population you are going to project, forecast, or estimate, the better.

3.3 Components of Change—Demographic Reasons for Population Change

Two simultaneous forces account for changes in population over time. First, population grows over time through births (B) and people moving into the target region (in-migrants; IM). Second, population declines through deaths (D) and people leaving the region (out-migrants; OM). Explaining observed and projected population changes through accounting for the individual components of change (births, deaths, and migration) sets the stage for probably the most basic formula in demography: the demographic balancing Eq. (3.3a) and Eq. (3.3b).

Demographic balancing equation is: [①]

$$P_{t+n} - P_t = B - D + (\mathrm{IM} - \mathrm{OM}) \tag{3.3a}$$

① Smith et al., 2001: 30.

The change in population between future year $t+n$ and initial year t is the result of the number of births (B) plus the in-migrants (IM), minus the number of deaths (D) and out-migrants (OM) for this specific time period of n years. The two population variables, P_{t+n} and P_t, are static measures and refer to the population statistics at one point in time. The four components of change measures are dynamic and quantify the number of births, deaths, and in- and out-migrants for that time period. Alternatively, in- and out-migrants could be netted out and referred to as "net migration (NM)". The demographic balancing equation can then be rewritten as:

$$P_{t+n} - P_t = B - D + \mathrm{NM} \tag{3.3b}$$

where NM refers to net migration which is computed as: $\mathrm{NM} = \mathrm{IN} - \mathrm{OM}$.

The population statistic P can be the total population. As we will explore in greater detail in the section on the cohort-component method, the population is most commonly further disaggregated into sex and age cohorts.

Taking a second look at the demographic balancing equation we can immediately recognize that part of population changes, e.g., B and D, come mainly from the population itself, independent from outside forces. This is where the sex and age structure of the population play a crucial role. More women in the childbearing age means more births per time period, everything else being constant. On the other hand, populations with an old age structure are more likely to have a greater number of deaths per time period. Netting out births and deaths, we can refer to **natural population increase or decrease**. In the case where births outnumber deaths ($B > D$), the population is said to experience natural increase. However, when more people are dying than being born ($B < D$), population would naturally decrease.

People migrate for various reasons. Some people just need a move, some are looking for better regional amenities (e.g., weather, recreational value of region), and others simply get transferred through their jobs. The Sunshine State, Florida, is the most popular retirement state in the United States. The weather is the main factor. Younger people seem more likely to prefer San Francisco, New York, or Washington D.C. for professional reasons, or simply because these places are the "happening" place to live. The bottom line is that reasons for migration are very complex. A structural population model is one strain of models that accounts for other than pure demographic factors included in the demographic balancing Eq. (3.3a).

3.3.1 Fertility

The first component on the right-hand side of the demographic balancing Eq. (3.3) deals with B, also referred to as fertility. Technically, the term fertility denotes the number of live births, often expressed as the number of actual live births to women in a particular age cohort, symbolized by *n*. Many different fertility rates are used in the literature. Table 3.4 lists two of them. The age-specific birth rate $({}_n\text{ASBR}_x)$ is the number of live births per 1,000 females in the same age cohort over an *x* year period. Fertility rates also can be presented as probabilities that a woman in this specific age cohort will give birth in *x* years period $P({}_n\text{ASBR}_x)$. In this case, we talk about fertility rates f_n, where *n* refers to the age cohort *n*.

Table 3.4 Live births and fertility rates, Boone County, Kentucky, July 2003(1)

Age-specific Cohort of Mother	Age-specific Birth Rates(1) $({}_n\text{ASBR}_x)$	Probability of Birth $(P({}_n\text{ASBR}_x))$
10 – 14	52.4	0.0524
15 – 19	391.6	0.3916
20 – 24	491.7	0.4917
25 – 29	491.7	0.4917
30 – 34	287.8	0.2878
35 – 39	86.4	0.0864
40 – 44	9.5	0.0095
Total		1.8111

(1) per thousand females over a five-year period

Table 3.4 represents the age-specific birth rates statistics for Boone County, Kentucky.① Following this table, we immediately can identify that:

(1) the number of live births are reported for a five year period.

(2) the females are divided into five-year age cohorts.

(3) the age range of childbirth-giving women starts at 10 – 14 years and ends at 40 – 44 years.

(4) the majority of childbirth-giving females are 20 – 29 years of age. Whereby in the case of Boone county, the lower half of the twenty age cohort (e.g., 20 – 24 years) and the upper half of the twenty age cohort (e.g., 25 – 29 years) appear to have identical fertility rates.

From Table 3.4, we can conclude that fertility rates are very age specific, which is graphically emphasized in Fig. 3.4.

① Source: Kentucky Population Projections, July 2003, The University of Louisville Urban Studies Institute, Kentucky State Data Center.

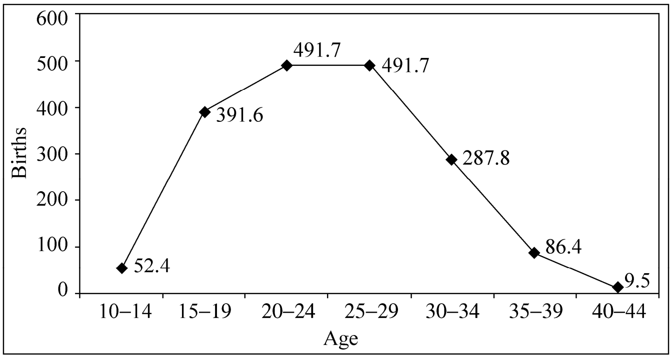

Figure 3.4 Age-specific live births, Boone County, Kentucky

There are many more factors that influence fertility rates. Among others, fertility rates vary geographically, culturally, and over time. According to the U.S. Central Intelligence Agency's (CIA, 2004) "The World Factbook", the United States' birth rate is estimated at 14.14 births/1,000 persons at midyear for the year 2003. For comparison, Ethiopia's birth rate for the same year is estimated at 39.81 births/1,000 persons and Germany's birth rate is at 8.6 births/1,000 persons indicating geographical and cultural variations in birth rates.

Looking at fluctuations of fertility rates over time, the United States exhibits some sharp changes. While the total fertility rate was slightly over 3.5 births per woman during the Baby Boom years in the late 1950s and early 1960s, it fell drastically by the mid-1970s to about 1.8 births per woman. Since the 1990s, the total fertility rate[①] per woman recovered slightly to a level of 2.0 – 2.1, which is the level required for the natural replacement of the population.[②]

While fertility rates differ across regions and fluctuate over time, it is important to recognize that these differences and fluctuations are not solely grounded in regional and cultural variations, but may be the result of a combination of complex economic, social, and other factors. This becomes apparent by comparing fertility rates across selected female subgroups for women 15 – 44 years old:[③]

(1) The general fertility rate (GFR) for all women was 61.4 births per 1,000 women.

(2) Hispanic women had the highest general fertility rate among all race and origin groups with 82.0 births per 1,000 women, while the GFR was significantly smaller for Asian and Pacific Islander women with 55.4.

① The total fertility rate refers to the average number of children that would be born per woman if all women lived to the end of their childbearing years and bore children according to a given fertility rate at each age.

② Source: U.S. Census Bureau, Fertility of American Women: June 2000 (P20-543RV). National Center for Health Statistics, National Vital Statistics Report, Vol. 47, No. 25.

③ All fertility rates are reported as births per 1,000 women for the year 2002. Source: Fertility of American Women: June 2002, October 2003, U.S. Census Bureau.

(3) Women in the South are slightly more reproductive than women living in the Midwest with fertility rates of 67.0 and 55.6 births per 1,000 women respectively.

(4) Women in the labor force, of course, had lower fertility rates (47.4) than women not in the labor force (95.0).

(5) Women with an annual family income of under $10,000 had the highest fertility rate (84.5) versus women with a family income of more than $75,000, who had the lowest fertility rate (60.0).

(6) With respect to educational attainments, the lowest fertility rate is reported for women who graduated from college with an associate degree (51.6), while women who received a graduate or professional degree have a significantly higher fertility rate (84.9).

We see that fertility rates vary across geographic regions, and are dependent on economic and social factors and fluctuate over time. We further see that different data sources, for instance the CIA World Factbook and the National Center for Health Statistics (NCHS), use different definitions of fertility/birth rates. They all build upon the total number of live births (reported in the United States by the NCHS) and the corresponding population size for that specific area (reported by the U.S. Census Bureau). In the conclusion of this section on births and fertility, we will define in more detail the different fertility rates.

Crude birth rate (CBR): The crude birth rate used in the example is from the CIA World Factbook as the number of live births per 1,000 population.

$$\text{CBR} = (B / P) \times 1{,}000 \tag{3.4}$$

where, B is the number of live births per year and P is the total midyear population.①

The National Center for Health Statistics announced on its 2003 National Vital Statistics Report (Vol. 52, No. 10) the lowest birth rate for the United States since national data have been available with a crude birth rate of 13.9. For the calculation, the NCHS used the reported live births during 2002 and the 2002 population estimate produced by the U.S. Census Bureau based on the 2000 census. Given all the necessary pieces of information, the crude birth rate formula② can be written as follows:

$$\text{CBR} = (4{,}021{,}726 / 288{,}368{,}706) \times 1{,}000 = 13.9$$

General fertility rate (GFR): The next logical step to improving the crude birth rate is by relating the number of births to the number of females in the

① Live births reports the number of babies born. Midyear population refers to the number of people alive at midyear, usually a calendar year.

② The number of births is taken from the National Vital Statistics Report, Vol. 52, No. 10, December 17, 2003: "Births: Final Data for 2002", p. 30, Table 1.

reproductive age group, mainly 15 – 44 years. The examples we used earlier to emphasize the influence of economic and social factors as determinants for differences in fertility rates used general fertility rates. The generic formula can be written as:

$$\text{GFR}=(B/\text{FP}_{15-44})\times 1{,}000 \tag{3.5}$$

where, B is the number of live births per year and FP_{15-44} is the age-specific female cohort of the population, or all women of age 15 – 44.

Using the GFR formula, the general fertility rate for all women of age 15 – 44 in the United States for the year 2002 can be computed as:[①]

$$\text{GFR} = (4{,}021{,}726 / 62{,}044{,}142) \times 1{,}000 = 64.8$$

Age specific birth rate (ASBR): The next step is to take all females in reproductive age for a specific area and divide them into age cohorts, for example, age cohorts of five years. This was done in Table 3.4 and the results are age-specific birth rates. For each age cohort of females, the ASBR is computed as the ratio of total births for that specific age group over the total number of females in this particular age group at midyear.

$${}_n\text{ASBR}_x = ({}_nB_x / {}_n\text{FP}_x) \times 1{,}000 \tag{3.6}$$

where x indicates the lower limit of the age cohort and n, the number of years in the age interval. B and FP refer, again, to the number of births and female population, respectively.

For example, ${}_n\text{FP}_x$ defines the age cohort that starts at age x and includes all females up to an age of $x+n$ years. ${}_5\text{FP}_{20}$ therefore, defines the cohort of females age 20 – 24, which spans five years, at midyear. Table 3.4 reports the ASBR for the cohort ${}_5\text{FP}_{20}$ to be 491.7. Meaning that for 1,000 women belonging to this age group, 491.7 will give birth over a five-year period.

Total Fertility Rate (TFR): Often, the literature refers to average number of births per woman. For example, earlier we reported that in the United States the average number of births per woman fluctuated between 2.0 and 2.1 during the 1990s. What people use here is called the total fertility rate. It is computed as the sum of all ASBR's. For example, Table 3.4 identifies the conditional probability—$P({}_n\text{ASBR}_x)$—that a woman will give birth given that she belongs to the age cohort 20 – 24 years with 0.4917. This probability refers to one woman

① Using births from the: National Vital Statistics Report, Vol. 50, No. 5 and the number of females of 15 – 44 years from the U.S. Census Summary File (SF1), the GFR is slightly different from the October 2001 release by the U.S. Census Bureau: GFR = 65.9 = (4,058,814 / 61,576,997) × 1,000. The difference is explained by the fact that at the earlier release date only preliminary data were available.

only and is conditional in that this woman must belong to that age-specific cohort. Accordingly, the total fertility rate for that specific woman would be adding all conditional probabilities for her entire reproductive lifespan:

$$\text{TFR}=\sum[P({}_n\text{ASBR}_x)] \tag{3.7}$$

From Table 3.4, we can calculate that the total fertility rate for Boone County, KY is 1.811. Although we know that the age-specific birth rates reported in Table 3.4 refer to the entire female population in Boone County at one point in time, namely 2003, for calculating the TFR, we now must interpret the table slightly different. Let us, for now, assume that we observe 1,000 women over their entire "hypothetical" reproductive lifespan. Meaning that in the beginning of their reproductive lifespan, all 1,000 women would be in the age group 10 – 14. Together they would give 52 live births during these five years. Accordingly, the same 1,000 women would after five years enter the second age cohort of 15 – 19 years and would give 392 live births. Analogously, adding all births together would tell us that over the reproductive lifespan of these specific 1,000 women, they would give 1,811 live births, assuming that no one left the cohort. In other words, the TFR refers to the number of babies born during women's reproductive years. In the case of Boone County, a woman on average will give birth to 1.8 babies in her life time.

Note that the TFR is based on hypothetical assumptions in that we now look at the entire lifespan of a group of women. We further assume that no one leaves this hypothetical cohort and the birth rates will not change over their lifespan. The TFR is also important when referring to the level necessary for natural replacement of the population; the replacement level fertility. According to the Census Bureau, approximately 2.1 births per woman are required for a population to maintain its current level in the long run. In the case of Boone County, KY, the TFR of 1.811 means that the population will naturally decline.

3.3.2 Mortality

A second component of change is people dying from one time period to the next. This can be either expressed in the form of **mortality or survival rates**. Table 3.5 lists the survival rates for Kentucky, as county-specific survival rates are not available for Kentucky.

The first column in Table 3.5 identifies the 5-year age cohorts. Columns two and three list the survival rates as the number of 5-year survivors per 1,000 male or females, respectively, by age cohort. Columns four and five express the same survival rates in form of probabilities for a person to survive from one age cohort to the next. For example, the entry for the female 25 – 29 age cohort of 996.6 states that of 1,000 females of age 25 – 29 (e.g., beginning age) 996.6 will

Table 3.5 Survival rates by age and sex, Kentucky, July 2003

Age-Specific Cohorts	Male	Female	Male	Female
Live births	992.4	993.4	0.9924	0.9934
0 – 4	996.7	997.5	0.9967	0.9975
5 – 9	998.8	999.2	0.9988	0.9992
10 – 14	997.2	998.5	0.9972	0.9985
15 – 19	992.9	997.3	0.9929	0.9973
20 – 24	992.9	997.4	0.9929	0.9974
25 – 29	991.8	996.6	0.9918	0.9966
30 – 34	989.6	995.3	0.9896	0.9953
35 – 39	985.9	993.2	0.9859	0.9932
40 – 44	980.3	989.1	0.9803	0.9891
45 – 49	970.9	983.5	0.9709	0.9835
50 – 54	954.0	973.0	0.9540	0.9730
55 – 59	925.6	957.0	0.9256	0.9570
60 – 64	883.0	932.0	0.8830	0.9320
65 – 69	823.6	897.0	0.8236	0.8970
70 – 74	750.8	850.4	0.7508	0.8504
75 – 79	639.2	765.3	0.6392	0.7653
80 – 84	498.5	637.3	0.4985	0.6373
85 +	297.6	381.4	0.2976	0.3814

survive the five-year period and enter the female age cohort of 30 – 35 (e.g., ending age). Expressed as a probability we can read the same entry in that there is a 0.9966 probability for a woman to reach age 30 assuming that she is in the 25 – 29 age cohort. The relationships of age, sex, and survival rates are graphically shown in Fig. 3.5. In Kentucky, females have higher survival rates at all ages. Both sexes show that with increasing age survival rates decline continuously. We also see immediately from the data that the first two age cohorts (e.g., 0 – 4 years and 5 – 9 years) show slightly lower survival rates which can be explained partly by higher infant and early childhood mortalities.

In the United States, the National Vital Statistics Report (NVSR) by the Department of Health and Human Services[①] reports annually life tables by age, race, and sex, which are the most detailed source available at present for survival rates. Among others, these tables report probabilities of dying between ages x and $x+1$, number of people surviving to age x, and number dying between ages x to $x+1$. Five-year survival rate tables can be derived from the life tables through aggregation of annual data into five-year age cohorts.

① http://www.cdc.gov/nchs/.

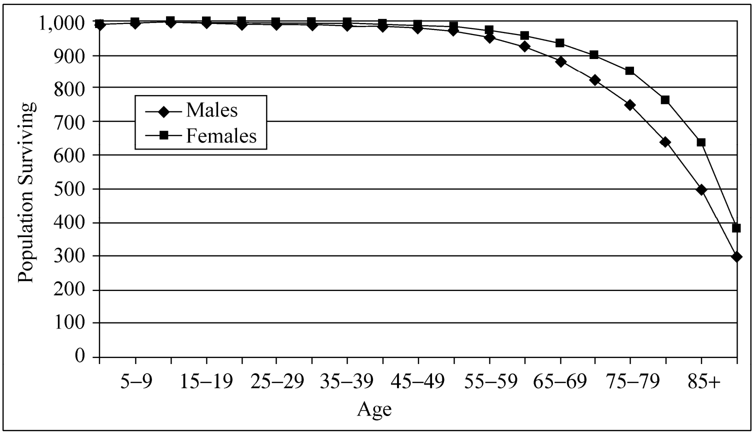

Figure 3.5 Population surviving by sex and age-specific cohort, Kentucky, July 2003

The NVSR also includes a table of survivorship by age, sex, and race in the United States from 1900 to 2000.[①] According to the table, the life expectancy rose significantly in this 100 year period. In 1900, 58.5% of the population reached age 50 and 13.5% survived to age 80. In 2000, 93.5% survived to age 50 and 51.0% of the population survived to age 80. However, while the first 50 years indicate a large reduction in infant mortality (infants born surviving the first year) from 87.6% in 1900 to 97.0% in 1950, the second half of the last century was characterized primarily by improvements in the surviving age of the older population.

Life expectancy in the United States is recorded in "The World Factbook" as 77.43 years for the total population in 2003. This refers to the average number of years to be lived by a group of people born in the same year, assuming constant mortality at each age in the future. For comparison, life expectancy in Ethiopia is 40.88 years, in Germany 78.54 years, and in China 71.96 years. These data are often used as a measure of overall quality of life in a country.

For the remainder of this section on mortality, we will focus exclusively on survival rates. It is important, however, to recognize that if we know the number of people surviving from one time period to the next, we immediately know the number of people dying for this specific time period and vice versa.

Life Table Survival Rates: Official life tables were introduced as early as the 1660s in London by the Englishman John Graunt[②]. In the United States they have been prepared since the beginning of 1900, first for every ten years and

① Source: National Vital Statistics Report, December 19, 2002, Vol. 51, No. 3, p. 38, Table 10. Survivorship by age, rage, and sex: Death registration States, 1900 – 1902 to 1919 – 1921, and United States, 1929 – 1931 to 2000; http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_03.pdf.

② Smith et al., 2001: 52.

since 1945 on an annual basis. Today, annual or decennial life tables are available at national and state-level. Table 3.6 shows an abridged version of the U.S. life table for the entire population.

Table 3.6 Abridged life table for the total population, United States, 2000(1)

Age	Probability of Dying Between Ages x to $x+n$ $({}_nq_x)$	Number Surviving to Age x (l_x)	Number Dying Between Ages x to $x+n$ $({}_nd_x)$	Person-Years Lived Between Ages x to $x+n$ $({}_nL_x)$	Total Number of Person-Years Lived Above Age x (T_x)	Expectation of Life at Age x (e_x)
0 – 1	0.00693	100,000	693	99,392	7,686,810	76.9
1 – 4	0.00131	99,307	130	396,916	7,587,418	76.4
5 – 9	0.00082	99,177	82	495,668	7,190,502	72.5
10 – 14	0.00104	99,095	103	495,278	6,694,833	67.6
15 – 29	0.00341	98,992	338	494,200	6,199,555	62.6
20 – 24	0.00479	98,654	473	492,113	5,705,355	57.8
25 – 29	0.00494	98,181	485	489,702	5,213,242	53.1
30 – 34	0.00578	97,696	565	487,130	4,723,539	48.3
35 – 39	0.00806	97,132	783	483,813	4,236,409	43.6
40 – 44	0.01182	96,349	1,139	479,070	3,752,596	38.9
45 – 49	0.01773	95,210	1,688	472,085	3,273,527	34.4
50 – 54	0.02576	93,522	2,409	461,940	2,801,442	30.0
55 – 59	0.03968	91,113	3,615	447,124	2,339,510	25.7
60 – 64	0.06133	87,498	5,366	424,879	1,892,377	21.6
65 – 69	0.09217	82,131	7,570	392,758	1,467,498	17.9
70 – 74	0.13838	74,561	10,317	348,168	1,074,739	14.4
75 – 79	0.20557	64,244	13,207	289,331	726,571	11.3
80 – 84	0.31503	51,037	16,078	215,947	437,240	8.6
85 – 89	0.46111	34,959	16,120	133,503	221,293	6.3
90 – 94	0.61506	18,839	11,587	62,766	87,790	4.7
95 – 99	0.75434	7,252	5,470	20,388	25,024	3.5
100 years and over	1.00000	1,781	1,781	4,636	4,636	2.6

(1) Source: National Vital Statistics Report, December 19, 2002, Vol. 51, No. 3, p. 38.

Column 1, Age-specific intervals (x to $x+n$): Reports the exact interval—n—between two ages—x and $x+n$—as indicated. For example, "5 – 9" indicates the five year interval between the fifth and tenth birthday. In this example, n indicates a five year interval and x equals the beginning age of the age cohort. Exceptions are the first two age cohorts age $0-1$ and $1-4$ and the last age cohort $100+$.

Column 2, Probability of dying between ages x to $x+n$ $({}_nq_x)$: Refers to the proportion of people alive at age x (beginning of the interval) and will not reach age $x+n$ (end of the interval). For instance, ${}_5q_{10} = 0.00104$ tells us that the proportion of the total population in the United States dying after their tenth birthday and before reaching their fifteenth birthday is 0.00104. Meaning that out of every 100,000 people in this cohort, 104 will die before reaching age fifteen.

Column 3, Number surviving to age x (I_x): Shows the number of the surviving members of 100,000 people at age x. Beginning with 100,000 life births in column three, 97,132 will complete their 35th year of life and 1,781 will have a 100 year birthday party.

Column 4, Number dying between ages x to $x+n$ $({}_nd_x)$: Reports the number of all the people dying from 100,000 life births between exact ages x to $x+n$. Following Table 3.6 we can identify that in the United States 693 babies will die in their first year of life and 13,207 people will die between ages 75 to 80.

Column 5, Person-years lived between ages x to $x+n$ $({}_nL_x)$: Refers to the total of person-years lived between ages x to $x+n$. Important for deriving person-years lived is knowing when people die during a particular age interval. For instance, people belonging to the age cohort $35-40$ who reach their 40th birthday would contribute five-person years each. People dying between ages 35 to 40 would contribute less than five person-years lived depending on when they died. In case they died exactly on their 37th birthday, they count as two person-years, if they died in between birthdays, they count for the whole years and partial years they lived.

Column 6, Total number of person-years lived above age x (T_x): Is the summation of the total of person-years lived between ages x to $x+n$ (e.g., ${}_nL_x$) and that of all subsequent age intervals. For instance, the aggregated total of person-years lived above age 35 (e.g., 4,236,409) is the sum of all intervals of person-years lived between ages x to $x+n$ (e.g. column 5) starting with interval 35 to 40.

Column 7, Expectation of life at age x (e_x): Indicates the remaining lifetime in years for persons reaching the exact age x. According to the abridged life table for the total population of the United States, persons of age 35 are expected to have an average remaining lifetime of 43.6 years.

The relationships among the different variables included in Table 3.6 and described thereafter can also be expressed in mathematical terms.

Column 2, the probability of dying between ages x to $x+n$ $({}_nq_x)$ can be calculated as:

$${}_nq_x = {}_nd_x / I_x$$

Knowing the number of people surviving between ages x to $x+n$, the number dying between ages x to $x+n$ can be expressed as:

$${}_nd_x = I_x - I_{x+n}$$

The person-years lived between ages x to $x+n$ $({}_nL_x)$ can be expressed as

$${}_nL_x = T_x - T_{x+n}$$

Data in a life table can then be used to calculate survival rates as:

$${}_nS_x = {}_nL_{x+n} / {}_nL_x \tag{3.8}$$

where ${}_nS_x$ is the survival rate, ${}_nL_{x+n}$ and ${}_nL_x$ are the numbers of two successive person-years lived for the corresponding successive age intervals taken from the life table.

For the total U.S. population, for example, the survival rate from the age interval 35 – 40 to 40 – 45 is computed as:

$${}_5S_{35} = {}_5L_{35+5} / {}_5L_{35} = 479{,}070 / 483{,}813 = 0.9902$$

When we plan to project population for male and female, we can find separate life tables for male and female and for different target areas. The structures of the life tables are exactly the same so is the calculation of the survival rates. Moving the decimal point three positions to the right, we now can compare this national level survival rate (e.g. 990.2) with the numbers reported for the same age interval in the Kentucky survival rate table. For males, the corresponding survival rate is reported as 989.6 and for females it is 995.3. Observed discrepancies in survival rates between the U.S. population and Kentucky are very small and can be explained through:

(1) different target years. The United States rates refer to the year 2000, while the Kentucky rates are for 2003.

(2) different levels of aggregation. We are comparing sex-specific rates (e.g., female and male in Kentucky) with aggregated rates for the entire U.S. population.

(3) different target areas. Here, we compare state-level with national-level survival rates.

Conceptually similar to the Age-Specific Birth Rates $({}_n\mathrm{ASBR}_x)$ are Age-Specific Death Rates $({}_n\mathrm{ASDR}_x)$ which play an important role as a starting point for the construction of life tables. They are calculated as the ratio of the number of deaths $({}_nD_x)$ between age x and age $x+n$ over the population $({}_nP_x)$. Subscript n refers to the time period of n years.

$$ {}_n\mathrm{ASDR}_x = {}_nD_x / {}_nP_x \tag{3.9} $$

Like fertility rates, ASDRs use the mid-interval population, preferably from census data. Mid-interval population is commonly used when averaging population. Alternatively, the population can refer to the number of people at the beginning of a period. For constructing life tables it is assumed that deaths are spread out evenly over the entire time-period, n.

3.3.3 Migration

The last two components of the demographic balancing equation account for the fact that people relocate. People move within the same county to a different residence, to a different county within the same state, between states, or even internationally. Generally, talking about moving or movers implies a change in location. People moving within a town from one end to the other end and staying within the same jurisdictional boundary are referred to as local movers and are not considered to be migrants. To qualify as a migrant, a person must move across jurisdictional boundaries. Although, for people living close to county or state boundaries, this can also imply just moving a few blocks away.

Depending on the attractiveness of a place, which among others includes amenities, availability of jobs, and recreational activities, etc., some places have positive net migration rates, indicating that more people move into the region than leave it. On the other hand, less attractive and declining regions show negative net migration rates. It is important to recognize that although net migration shows whether a region gains or loses population due to migration, it does not implicitly indicate how many people are actually migrating in and out. Depending on the depth of your analysis and data availability, you might want to account for in- and out-migration separately or simply use net migration data.[①] The U.S. Census Bureau provides data on migration in the United States at national, state and county levels.[②] Table 3.7 shows that Americans indeed are still on the move. Forty-six percent of Americans ages 5 years and over,

---

① In the concept of planning in general and in population projection for planning purposes in particular, net migration is sufficient most of the time. For a more detailed discussion on the pros and cons of gross versus net migration please see chapter 6 of Smith, Tayman, and Swanson.

② Data Source: Census 2000 PHC-T-23. Migration by Sex and Age for the Population 5 Years and Over for the United States, Regions, States, and Puerto Rico: 2000 at: http://www.census. gov/population/www/cen2000/migration.html.

Table 3.7 Gross and net migration by sex for the population 5 years and over for the United States and Kentucky, 2000

Geographic Area	Population 5 years and over	Same residence (nonmovers)	Different residence 5 years ago					
			Total movers	Different residence in same geographic area	Domestic Migration[1]			From abroad[2]
					Inmigrants	Outmigrants	5-year net migration	
United States	262,375,152	142,027,478	120,347,674	112,851,828	(X)	(X)	(X)	7,495,846
Male	128,160,479	68,381,473	59,779,006	55,767,900	(X)	(X)	(X)	4,011,106
Female	134,214,673	73,646,005	60,568,668	57,083,928	(X)	(X)	(X)	3,484,740
Kentucky	3,776,230	2,112,135	1,664,095	1,299,535	318,579	284,452	34,127	45,981
Male	1,838,610	1,019,192	819,418	629,222	163,929	144,452	19,477	26,267
Female	1,937,620	1,092,943	844,677	670,313	154,650	140,000	14,650	19,714

(1) Outmigrants and 5-year net migration are not included in total movers.

(2) This category includes movers from foreign countries, as well as movers from Puerto Rico, U.S. Island Areas, and U.S. minor outlying islands.

or 120,347,674 people, moved in the period from 1995 to 2000. Of these 120 million movers, only a small portion (7,495,846 people or approximately 6 percent) qualify as migrants. At the state-level for Kentucky, for example, 44% (1,664,095 people) of Kentucky's population moved in the same five-year period, of which 318,579 moved into the state and 284,452 left the state resulting in net migration of 34,127 people. Additionally, 45,981 people moved in from abroad.

While calculating fertility and survival rates is usually done based on area-specific data, for computing in/out and net migration rates, the choice of the appropriate population base is not that straightforward. In the case of out-migration, we can apply the areas of interest population as base but what can be applied in the case of in-migration rates? Definitely, these rates are somewhat independent of the target area. Of course, some areas, like San Francisco, are attractive enough to attract people. But in most cases, in-migration rates depend to a larger extent on the population statistics from where these people are migrating. This problem is well recognized in the literature; nevertheless, most studies applied for simplicity the population of the area under consideration as the denominator without differentiating between in/out and/or net migration rates.[①]

The general net migration rate is of the following form:

$$ {}_n\mathrm{mr}_x = ({}_nM_x / P_x) \times 1{,}000 \tag{3.10} $$

where ${}_n\mathrm{mr}_x$ is the migration rate under consideration (e.g., in, out, or net migration rate), ${}_nM_x$ is the corresponding number of in/out or net migrants between time period x to $x+n$, and P_x is the population total in beginning year x. In the case of calculating migration rates, the population total in beginning year x is preferred in the literature versus using midyear or end of interval population.

For Kentucky and Boone County, we can apply the migration rate formula and compute all rates as follows: [②]

The results in Table 3.8 show that Boone County has significantly larger migration rates than Kentucky. Kentucky experienced, from 1995 – 2000, a population increase due to net migration of almost 9 persons per 1,000 population, while Boone County's population grew by 119 persons per 1,000 population. A direct and logical conclusion from observed magnitudes of migration rates is that in the case of Boone County, in-migration plays a crucial role in population growth. In the case of Kentucky, which exhibits a slowly growing population, in-migration is a less significant factor of population growth.

---

① Smith et al., 2001: 105.

② The U.S. Census Bureau used in its 1995 – 2000 migration rate estimation the 1995 – 2000 net migration as the numerator and the approximated 1995 population as the denominator. Multiplying this rate by 1,000 then refers to rates per 1,000 population. Source: http://www.census.gov/prod/2003pubs/censr-7.pdf.

Table 3.8 Migration rate example, Kentucky and Boone County, 1995 – 2000[1]

Kentucky	Number of Persons	Rates (per 1,000 Population)
Total Population (1995)	3,887,427	
Total Population (2000)	4,041,769	
Inmigrants (1995 – 2000)	318,579	81.95
Outmigrants (1995 – 2000)	284,452	73.17
Gross Migration (1995 – 2000)	603,031	155.12
Net Migration (1995 – 2000)	34,127	8.78
Boone County	**Number of Persons**	**Rates (per 1,000 Population)**
Total Population (1995)	70,017	
Total Population (2000)	85,991	
Inmigrants (1995 – 2000)	25,232	360.37
Outmigrants (1995 – 2000)	16,896	241.31
Gross Migration (1995 – 2000)	42,128	601.68
Net Migration (1995 – 2000)	8,336	119.06

(1) Sources: http://ksdc.louisville.edu/kpr/pro/Summary_Table.xls and http://www.census.gov/prod/2003pubs/censr7.pdf.

Referring back to the problem of the correct population choice for the denominator in computing the migration rates, using the population under consideration will have different effects in both cases. Generally, the literature recommends using the population under consideration as the denominator for decreasing or slowly growing populations. Thus in the case of Kentucky, using Kentucky's population will not introduce a large error in computing migration rates. However, in the case of Boone County, the argument might be that in-migration by far outweighs other migration forces, and that this in-migration is independent of the target areas population. Therefore, a situation where choosing Kentucky's or even the United State's population might be the better denominator for calculating migration rates. For example, the net migration rate for Boone County can be recalculated using Kentucky's population as denominator:

$$\text{net migration rate} = (8{,}336 \,/\, 3{,}887{,}427) \times 1{,}000 = 2.14$$

This example shows that by using Kentucky's larger population, we get a significantly smaller net migration rate. Overall, using Kentucky's population will help smooth out the otherwise significant impact of a large number of in-migrants which migrated, to a large extent, independent of Boone County's existing population. But keep in mind, when we are projecting future net migration in Boone County, we must use the population of Kentucky as the basis for calculating the number of in migrants.

3.4 Trend Extrapolation Methods

Extrapolating past trends into the future is the main idea behind all the trend extrapolation methods. This idea is appealing for small-area population projections with low data requirements, low costs, and easy application. Observing how the population grew/declined for the past years, we project future population assuming that these observed trends will continue into the near future. For example, Boone County's, Kentucky population increased in each consecutive year between 1990 and 2000 (Table 3.9). Continuing past trends into the future will, therefore, lead to projecting the population as growing in the future.

Table 3.9 Population of Boone County, Kentucky, 1990 – 2000[1]

Year (n)	Index Number	Observed Population (Pop_n)	Total Absolute Growth	Average Annual Absolute Change(AAAC)	Total Percent Growth	Average Annual Percent Change (AAPC)
1990	1	57,589				
1991	2	60,574	28,402	2,840	49.32	4.09
1992	3	62,897				
1993	4	65,318				
1994	5	67,554				
1995	6	70,017				
1996	7	72,860				
1997	8	76,162				
1998	9	79,818				
1999	10	83,349				
2000	11	85,991				

(1) Source of data: http://ksdc.louisville.edu/kpr/popest/ice9000.xls.

A fast and straight forward approach to getting a first impression on the overall past population trend can be obtained here by plotting the population on a simple graph with time on the horizontal axis and population on the vertical axis. Although population projection(s) are usually based on a mathematical model, it is particulary the graphical presentation the helps better understand population trends over time. The case of Boone County is clear cut in that the population grew continuously in the past as indicated in Fig. 3.6. The trend could be more complex in that population both increased and decreased during the period in which data are available.

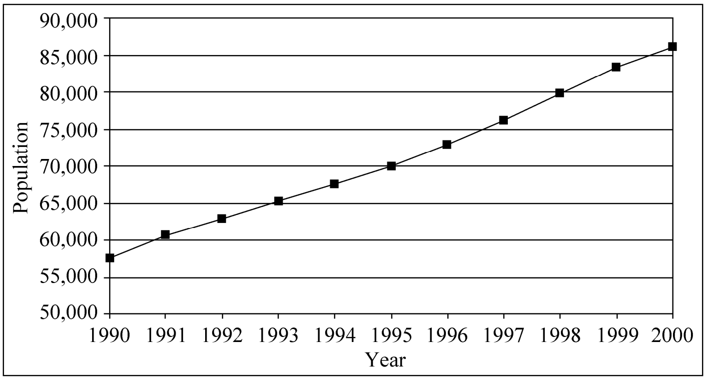

Figure 3.6 Population of Boone County, Kentucky, 1990 – 2000

We start describing population extrapolation models with simple average annual absolute population changes (AAAC) and/or average annual percent changes (AAPC) based on population statistics at two points in time. For instance, having the census population statistics for an area for the Census 1990 (Boone County: 57,589) and the Census 2000 (Boone County: 85,991) is sufficient to immediately compute two estimates: (1) the estimate for the observed annual population growth expressed as persons per year (AAAC) and (2) the estimate for the constant annual rate the population grew over the time period for which data are available (AAPC). For Boone County, the AAAC is derived by dividing the absolute population growth by the number of years:

$$\begin{aligned}\text{AAAC} &= (\text{Pop}_{2000} - \text{Pop}_{1990})/n \\ &= (85{,}991 - 57{,}589)/10 \\ &= 2{,}840 \end{aligned} \tag{3.11}$$

Based on the same information, the AAPC is calculated using the geometric growth rate formula as:

$$\begin{aligned}\text{AAPC} &= (\text{Pop}_{2000} / \text{Pop}_{1990})^{(1/n)} - 1 \\ &= (85{,}991 / 57{,}589)^{(1/10)} - 1 \\ &= 4.09\% \end{aligned} \tag{3.12}$$

With this information readily available, we can have a quick and simple population projection for 2001, assuming that the observed average annual absolute/percent changes will continue for the following year. More specifically, applying the average annual absolute change, the population for 2001 is projected as:

$$\begin{aligned}\text{Pop}_{2001} &= \text{Pop}_{2000} + n \cdot (\text{AAAC}) \\ &= 85{,}991 + 1 \times (2{,}840) \\ &= 88{,}831 \end{aligned}$$

where, Pop refers to the population in the corresponding years, *n* is the number of years to project in the future, and AAAC is the average annual absolute change of the area's population. Overall, the population follows a linear growth pattern, depending solely on the calculated average annual absolute change for the period data are available.

Analogously, we can apply the already computed average annual percent change (AAPC) rate as follows:

$$\begin{aligned} \mathrm{Pop}_{2001} &= \mathrm{Pop}_{2000} \cdot (1 + \mathrm{AACP})^n \\ &= 85{,}991 \times (1 + 0.0409)^1 \\ &= 89{,}508 \end{aligned}$$

where, Pop again refers to the population in the corresponding years, *n* is the number of years to project in the future, and AAPC is the growth rate expressing average annual percent changes of the area's population. what is important here is that the population grows annually by the same rate, namely by 4.09 percent. The difference in projected Boone County population for the years 2001 – 2010 using the AAAC and the AAPC is shown graphically in Fig. 3.7.

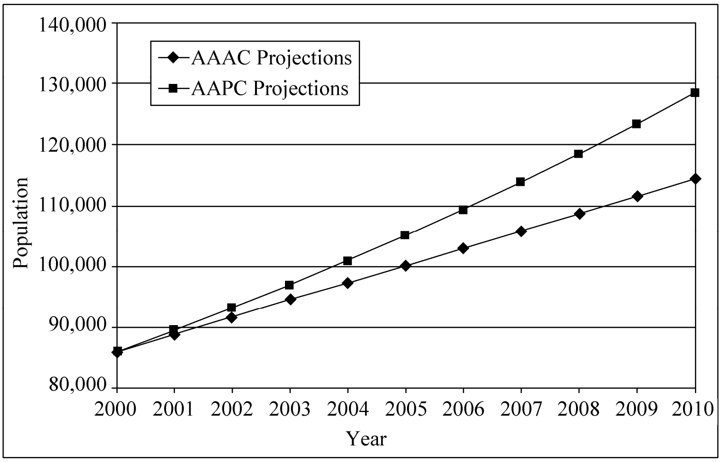

Figure 3.7 Boone County population projections based on average annual absolute change and average annual percent change, 2001 – 2010

3.4.1 Share of Growth Method

Ratio methods, such as the share of growth and the shift-share methods, are among the easiest extrapolation methods and are therefore popular among planners and demographers. The underlying principle of the share of growth as well as the shift-share method is a comparison of the smaller area's population to the population of a larger area, such as comparing a county to a state or a metropolitan area. In particular, the share of growth method observes the smaller area's share of population growth for a past time period—the base period. Assuming that this observed share of growth remains constant and knowing the larger area's projected population for the future target year, we can project the smaller area's future population. The share of growth method is expressed as:

$$
\begin{aligned}
\mathrm{Pop}_{m,\mathrm{ty}} &= \mathrm{Pop}_{m,\mathrm{ly}} + \left[\frac{(\mathrm{Pop}_{m,\mathrm{ly}} - \mathrm{Pop}_{m,\mathrm{by}})}{(\mathrm{Pop}_{n,\mathrm{ly}} - Pop_{n,\mathrm{by}})}\right](\mathrm{Pop}_{n,\mathrm{ty}} - \mathrm{Pop}_{n,\mathrm{ly}}) \\
&= \mathrm{Pop}_{m,\mathrm{ly}} + \mathrm{growthshare}(\mathrm{Pop}_{n,\mathrm{ty}} - \mathrm{Pop}_{n,\mathrm{ly}}) \qquad (3.13)
\end{aligned}
$$

where,

Pop_m — population of smaller area;

Pop_n — population of larger comparison region;

ty — target year, i.e., year to be projected;

ly — launch year, i.e., later year of base period;

by — base year, i.e., earlier year of base period;

growthshare — share of growth.

An example of Boone County demonstrates the share of growth method. The needed data are listed in Table 3.10.

Table 3.10 Boone County and Kentucky population statistics, 1990 – 2000(1)

Year	Kentucky	Boone County
1990	3,686,891	57,589
2000	4,041,769	85,991
2010	4,374,591	

(1) Source: Kentucky State Data Center, Summary table for Kentucky and Counties: http://ksdc.louisville.edu/kpr/pro/Summary_Table.xls.

The base period in the example is the period from 1990 to 2000. The observed share of growth for this ten-year period is calculated as:

$$
\mathrm{growthshare} = \frac{\mathrm{Pop}_{\mathrm{Boone},2000} - \mathrm{Pop}_{\mathrm{Boone},1990}}{\mathrm{Pop}_{\mathrm{KY},2000} - \mathrm{Pop}_{\mathrm{KY},1990}} = \frac{85,991 - 57,589}{4,041,769 - 3,686,891} = 0.08
$$

Assuming this share of growth of 0.08 for Boone County to remain constant in the future and knowing Kentucky's population for 2010, we can project Boone County's population for the year 2010 as:

$$\begin{aligned}\text{Pop}_{\text{Boone,2010}} &= \text{Pop}_{\text{Boone,2000}} + \text{growthshare} \cdot (\text{Pop}_{\text{KY,2010}} - \text{Pop}_{\text{KY,2000}}) \\ &= 85{,}991 + 0.08 \times (4{,}374{,}591 - 4{,}041{,}769) \\ &= 85{,}991 + 26{,}626 \\ &= 112{,}617\end{aligned}$$

Although the share of growth method is very simple in its application, there are situations where the share of growth method cannot be applied. Imagine a situation where, for instance, a county with a declining population is situated in an otherwise growing state. If we predict the population for the larger area to increase faster than previously observed, then the share of growth would predict the smaller area to decline faster as observed for the projection period. One can justly argue that this is a very unlikely assumption. The share of growth method must be applied with care in cases where smaller and larger areas' populations are not moving in the same direction.

3.4.2 Shift-Share Method

Rather than using shares of growth, the shift-share method uses the smaller area's share of total population in the base year and in the launch year. These two population shares and the projected population for the larger comparison region for the target provide the means for applying the shift-share method as:

$$\begin{aligned}\text{Pop}_{m,\text{ty}} &= \text{Pop}_{n,\text{ty}}\left[\frac{\text{Pop}_{m,\text{ly}}}{\text{Pop}_{n,\text{ly}}} + \left(\frac{\text{years}_{\text{pp}}}{\text{years}_{\text{bp}}}\right)\left(\frac{\text{Pop}_{m,\text{ly}}}{\text{Pop}_{n,\text{ly}}} - \frac{\text{Pop}_{m,\text{by}}}{\text{Pop}_{n,\text{by}}}\right)\right] \\ &= \text{Pop}_{n,\text{ty}}\left[\text{share}_{\text{ly}} + \left(\frac{\text{years}_{\text{pp}}}{\text{years}_{\text{bp}}}\right)(\text{share}_{\text{ly}} - \text{share}_{\text{by}})\right] \qquad (3.14)\end{aligned}$$

where,

Pop_m — population of smaller area;

Pop_n — population of larger comparison region;

ty — target year, i.e., year to be projected;

ly — launch year, i.e., later year of base period;

by — base year, i.e., earlier year of base period;

share_{ly} — population share in launch year;

share_{ly} — population share in base year;

years_{pp} — number of years in the projection period;

years_{bp} — number of years in the base period.

Using again the population data from Table 3.10, we can project Boone County's population for the year 2010 using the shift-share method as:

$$\begin{aligned}
\text{Pop}_{\text{Boone,2010}} &= \text{Pop}_{\text{KY,2010}}\left[\frac{\text{Pop}_{\text{Boone,2000}}}{\text{Pop}_{\text{KY,2000}}}+\left(\frac{\text{years}_{2000-2010}}{\text{years}_{1990-2000}}\right)\left(\frac{\text{Pop}_{\text{Boone,2000}}}{\text{Pop}_{\text{KY,2000}}}-\frac{\text{Pop}_{\text{Boone,1990}}}{\text{Pop}_{\text{KY,1990}}}\right)\right]\\
&= 4,374,591\times\left[\frac{85,991}{4,041,769}+\left(\frac{10}{10}\right)\times\left(\frac{85,991}{4,041,769}-\frac{57,589}{3,686,891}\right)\right]\\
&= 4,374,591\times\left[0.02128+\left(\frac{10}{10}\right)\times(0.02128-0.01562)\right]\\
&= 117,851
\end{aligned}$$

The example of the shift-share method assumes linearly changing shares for the projection period. Alternatively, the population shares can follow a nonlinear growth pattern over time. Notable of the shift-share method is also that the last term in parenthesis, i.e. the shift-term, can be negative. This is always the case where the population shares of the smaller region declined over the base period. One implication of declining population shares is that for particularly long projection periods, the smaller area's population projection can turn out to be negative, which is not possible. We must evaluate the projected population with caution. Often, a comparison of the outcome of the population projection with the small area's population growth/decline for the base period can give some first clues as to whether or not the outcome of the population projections using ratio methods (e.g., share of growth and shift-share method) leads to reasonable results. Here, a good knowledge of the small and larger areas' past and present population trends will be a useful guide for interpreting the projection results.

The remainder of this section on extrapolation methods deals with more complex population models that use regression analysis to project future population trends. We use a hypothetical example to demonstrate the rationale behind regression analysis for population projections. We then introduce four different population models: (1) the linear population model, (2) the geometric population model, (3) the parabolic population model and (4) the logistic population model.

3.4.3 Linear Population Model

In addition to simple ratio methods described above, trend extrapolation models can use regression analysis to fit a line to observed population data. Because of its computational and conceptual ease, the linear population model as expressed in Eq. (3.15) is the most widely used population model:[1]

[1] Klosterman, 1990: 9; Smith et al., 2001: 167.

$$\mathrm{Pop}_n = \alpha + \beta \cdot T_n \tag{3.15}$$

where,

Pop_n — estimated population for a given year n;

α — intercept of the linear regression model;

β — slope coefficient of the linear regression model;

T_n — the index number for year n.

The main assumption on which the linear model is based is straight forward: the population growth follows a linear pattern, meaning that the population will grow by the same number of people every consecutive year, expressed by the slope, β. The graphic solution is represented by fitting a straight line as "closely as possible" to observed population data, as indicated in Fig. 3.8. Using the calculated linear trend line, future population projections will then be exactly on the line.

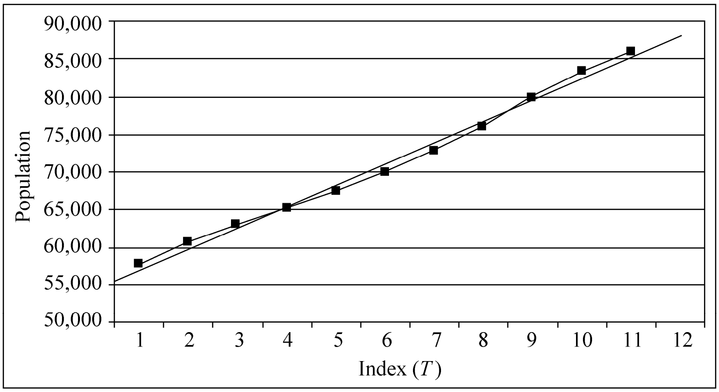

Figure 3.8 The linear trend line for Boone County, KY population data

As Fig. 3.8 indicates, the fitted straight line is an approximation of the observed population, but none of the observed data points (■) may actually lie on the straight line. The regression line has been fitted to observed population data for Boone County following the "least square criterion". As common in time series analysis, we supplemented the actual years (n), 1990, 1991, $\cdots$, 2000 with index numbers (T), e.g., 1, 2, $\cdots$, 10, 11, to simplify the computational process of estimating the regression line for the eleven years of available data for Boone County.

For a linear population trend line, the slope β indicates the calculated annual absolute population growth. In other words, it determines the number of people by which population grows/declines annually. Therefore, to determine a particular linear population trend line, we need to calculate the parameters α and β, which we will discuss now in greater detail. To demonstrate the computational process by hand, we are using Boone County's population statistics and the following linear regression model:

$$P = \alpha + \beta \cdot T + \varepsilon \tag{3.16}$$

The computational steps deriving the two parameters α and β are rather straight forward and outlined in detail in Table 3.11.

Table 3.11 Linear population trend line computations, Boone County, KY(1)

	Original Data			Deviations from Mean Values		Necessary Cross-products	
	Observed Population, Pop	Year, n	Index Numbers, T	Population, p	Index Number, t	$p \cdot t$	t^2
	57,589	1990	1	– 13,514	– 5	67,568	25
	60,574	1991	2	– 10,529	– 4	42,115	16
	62,897	1992	3	– 8,206	– 3	24,617	9
	65,318	1993	4	– 5,785	– 2	11,569	4
	67,554	1994	5	– 3,549	– 1	3,549	1
	70,017	1995	6	– 1,086	0	0	0
	72,860	1996	7	1,757	1	1,757	1
	76,162	1997	8	5,059	2	10,119	4
	79,818	1998	9	8,715	3	26,146	9
	83,349	1999	10	12,246	4	48,985	16
	85,991	2000	11	14,888	5	74,442	25
Totals	782,129		66	4	0	310,867	110

(1) Henceforth we adopt the convention of letting the lowercase letters p and t denote deviations from mean values for population statistics and index numbers.

(2) Population Mean: 71,103

(3) Index Number Mean: 6

As Table 3.11 shows, the table can be broken down into three distinct sections:

(1) the first three columns, including the original observed population data, the corresponding years (1990 – 2000), and the index numbers (1 – 11).

(2) the two successive columns, four and five, containing deviations from the mean values for the population (Pop) and the index number (T). Note that these deviations are denoted using lower case letters, e.g., p and t. Column four contains the deviations of the individual population statistics from the population mean. For example, – 13,514 is the difference between number of people in

Boone County in year 1 and the mean population for the 11 years, e.g., $(57{,}589 - 71{,}103) = -13{,}514$. Analogously, column five contains the deviations for the index number from their mean value, e.g., 6. For instance, the first deviation is computed as: $1 - 6 = -5$. To control the correctness of your computations, the sum of all deviations from the means must always equal zero.

(3) the last two columns, column six and seven, involve taking cross-products. Column six is the cross-product of the population deviation times the index number deviation (e.g., $p \cdot t$). For instance, 67,568 is the product of $-13{,}514$ and -5. The last column, seven, is the squared index number deviations (e.g., t^2). For instance, the first value 25 is $(-5)^2$.

What remains is plugging the results from Table 3.11 into the intercept and slope coefficient formulas:

$$\hat{\beta} = \frac{\sum(\text{Pop} - \overline{\text{Pop}}) \cdot (T - \bar{T})}{\sum(T - \bar{T})^2} = \frac{\sum p \cdot t}{\sum t^2} \tag{3.17}$$

where,

$\overline{\text{Pop}}$ — the mean of Pop;

$\bar{T}$ — the mean of T;

p — the deviations from $\overline{\text{Pop}}$;

t — the deviations from $\bar{T}$;

$\sum$ — the summation expression (e.g., column total).

The intercept formula uses the fact that the straight line passes through computed mean values $\overline{\text{Pop}}$ and $\bar{T}$:

$$\hat{\alpha} = \overline{\text{Pop}} - \hat{\beta} \cdot \bar{T} \tag{3.18}$$

For Boone County, the slope coefficient is estimated as

$$\hat{\beta} = \frac{\sum p \cdot t}{\sum t^2} = \frac{310{,}867}{110} = 2{,}826$$

The intercept $\hat{\alpha}$ can be computed using the estimated slope parameter, $\hat{\beta}$, and the mean values:

$$\hat{\alpha} = \overline{\text{Pop}} - \hat{\beta} \cdot \bar{T} = 71{,}103 - 2{,}826 \times 6 = 54{,}146$$

Inserting the parameter results into the linear population model for Boone County gives us the estimated population model:

$$\text{Pop} = 54{,}146 + 2{,}826 \cdot T$$

Given that the estimated population model has been derived by using index numbers, for projecting Boone County's population for future years we have to use index numbers as well. For instance, we use 12 for the year 2001, 13 for 2002, 14 for 2003, and so on. Using the appropriate index numbers and the estimated linear population model, Boone County's population for the year 2001, 2002, and 2003 can be computed as follows:

$$\hat{\text{P}}\text{op}_{2001} = 54{,}146 + 2{,}826 \times 12 = 88{,}058$$

$$\hat{\text{P}}\text{op}_{2002} = 54{,}146 + 2{,}826 \times 13 = 90{,}884$$

$$\hat{\text{P}}\text{op}_{2003} = 54{,}146 + 2{,}826 \times 14 = 93{,}710$$

Of course, the projection period can be extended further into the future with the assumption that a linearly growing population remains constant. However, future projections must be interpreted with caution, as they usually become more and more unreliable. This is due to the fact that the assumption of a linear growth rate might not hold over an extended period of time.

Using the estimated linear population model we can also calculate the population for years for which we have observed population data. For instance, we can calculate the population for the year 2000 as:

$$\hat{\text{P}}\text{op}_{2000} = 54{,}146 + 2{,}826 \times 11 = 85{,}232$$

Comparing the estimated population value of 85,232 with the observed population value of 85,991 for the year 2000, we notice that the estimated value underestimates the observed population by 759 people. Referring back to Fig. 3.8 this also can be seen in that the fitted regression line lies below the actual population for the year 2000. To achieve consistency of the estimated population value and observed population value for the last year population data are available—the launch year—Smith et al. (2001) recommend the inclusion of an ***adjustment factor***. They calculate the adjustment factor (ADJUST) as the difference of estimated and observed population for the launch year, or:

$$\text{ADJUST} = \text{observed Pop}_{2000} - \text{estimated Pop}_{2000}$$

Adding the adjustment factor to the linear population model for Boone County then shifts the fitted regression line upwards by 759 people and thus achieves consistency of estimated and observed population for the last year data were available. The general adjusted linear population model is:

$$\text{Pop} = \alpha + \beta \cdot T + \text{ADJUST} \tag{3.19}$$

And, for Boone County, the adjusted linear population model is defined as

$$\text{Pop} = 54{,}146 + 2{,}826\ T + 759$$

Using the adjusted linear population model, Boone County's population for years 2001, 2002, and 2003 can be recalculated as follows:

$$\widehat{\text{Pop}}_{2001} = 54{,}146 + 2{,}826 \times 12 + 759 = 88{,}817$$
$$\widehat{\text{Pop}}_{2002} = 54{,}146 + 2{,}826 \times 13 + 759 = 91{,}643$$
$$\widehat{\text{Pop}}_{2003} = 54{,}146 + 2{,}826 \times 14 + 759 = 94{,}469$$

The sole purpose of the adjustment factor is to match estimated and observed population for the last year data were available. Particularly in cases where the estimated population for the launch year differs significantly from the observed population, the exclusion of the adjustment factor can lead to unjustifiable overestimated/underestimated populations for the years to come. We cannot say with certainty whether or not the inclusion of the adjustment factor will improve the population projections after all.

3.4.4 Geometric Population Model

In many cases population data do not exhibit linear growth patterns when plotted on a simple scatter plot. Under certain circumstances, populations might grow or decline following constant growth rates. For example, a population might grow/ decline at a rate of approximately 4%. Note the difference, while linear models assume constant absolute population growth, for instance the population will grow incrementally by 2,500 people per time period, geometric population models assume the population grows/declines at a constant growth rate, expressed in percent. For instance, a population of 1,000 growing at a rate of 10% will grow by 100 people between year one and two, by 110 people between year two and three and by 121 people between year three and four. As you easily can see, the rate of growth remains constant at 10%, but the absolute numerical value increases every year by 100, 110 and 121 in the first three years. The direct conclusion of applying constant growth rates is that population grows slowly in earlier years but grows considerably faster in later years.

Using the simplified example, the growth rate, r, is defined as:

$$r = \frac{\text{Pop}_{n+1} - \text{Pop}_n}{\text{Pop}_n} = \frac{121 - 110}{110} = 0.1 \text{ or } 10\%$$

where:

r — constant growth rate;

Pop_n — population in year n;

Pop_{n+1} — population in year $n+1$.

The constant growth rate, r measures the rate of growth between year n and year $n+1$.

We now extend the simple example from before to a total of 19 years, use an initial population of 1,000, a constant growth rate of 10%, and plot the linear versus the geometric curve. The constant incremental growth for the linear line is set at 200 people per year. Figure 3.9 graphically shows the difference between a linear and a geometric population growth curve. We see that the linear model predicts a higher population until year 16. The population projected with a geometric model exceeds the linear model projection after year 16. The more years we would move to the right side of the graph, the larger the difference between linear and geometric model estimates would become.

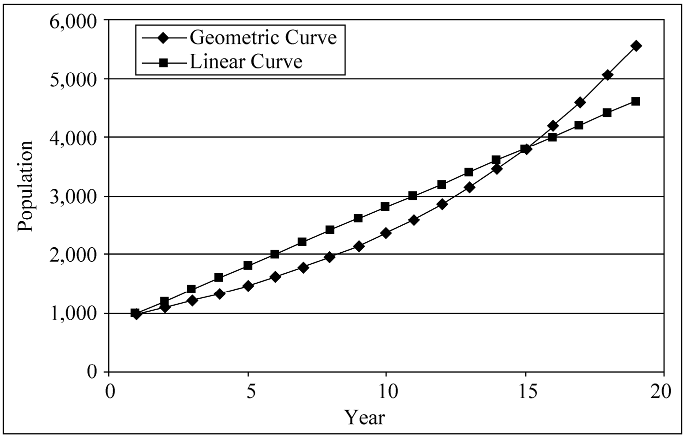

Figure 3.9 Linear line versus geometric curve

The compound rate formula

The growth rate is also commonly known as the compound rate. Here, the generic compound formula is:

$$\mathrm{FV}_n = \mathrm{PV} \cdot (1+i)^n$$

where,

FV_n — future value;

PV — present value;

i — compound interest rate;

n — number of years, e.g., time.

Through substitution of FV_n with Pop_n (i.e., projected population in year n), PV with α (e.g., initial population), $(1+i)$ with β (e.g., constant population growth factor), and n with T_n (the index number for year n) the compound rate formula takes the form of the geometric population model discussed in this section.

The general equation for the geometric growth curve is:

$$\text{Pop}_n = \alpha \cdot \beta^{T_n} \tag{3.20}$$

where,

Pop_n — population in year n;

T_n — index number for year n;

β — constant population growth factor;

α — initial population.

While the estimation procedure was relatively straight forward in the case of the linear model, the geometric population model requires one additional step to be able to use the "ordinary least-square" criterion to estimate the two parameters. The geometric equation needs to be transformed into a linear form. This is done by taking logarithms:

$$\text{Pop}_n = \alpha\beta^{T_n}$$
$$\log(\text{Pop}_n) = \log(\alpha\beta^{T_n})$$
$$\log(\text{Pop}_n) = \log(\alpha) + \log(\beta) \cdot T_n$$

where,

$\log(\beta)$ — slope of the population trend line in logarithmic form;

$\log(\alpha)$ — intercept of the trend line with the y-axis in logarithmic form;

$\log(\text{Pop}_n)$ — log value of the predicted population.

Table 3.12 below shows the necessary steps to estimate the two regression parameters using the "ordinary least-square" criterion. The sole difference to the

Table 3.12 Geometric population curve computations, Boone County, KY

	Observed Population, Pop	Logarithm of Obs. Population, log (Pop)	Index Numbers, T
	57,589	4.7603	1
	60,574	4.7823	2
	62,897	4.7986	3
	65,318	4.8150	4
	67,554	4.8297	5
	70,017	4.8452	6
	72,860	4.8625	7
	76,162	4.8817	8
	79,818	4.9021	9
	83,349	4.9209	10
	85,991	4.9345	11
Totals	782,129	53.3328	66

linear model is now that we have to take the logarithmic of observed population values as indicated in the second column.

The population mean, $\overline{\text{Pop}} = 71{,}103$ and the mean index, $\overline{T}$=6.

	Deviations from Mean Values		Necessary Cross-products	
	Population, $\log(p)$	Index Number, t	$\log(p)\cdot t$	t^2
	– 0.0881	– 5	0.4405	25
	– 0.0662	– 4	0.2646	16
	– 0.0498	– 3	0.1494	9
	– 0.0334	– 2	0.0668	4
	– 0.0188	– 1	0.0188	1
	– 0.0032	0	0.0000	0
	0.0141	1	0.0141	1
	0.0333	2	0.0666	4
	0.0537	3	0.1610	9
	0.0725	4	0.2898	16
	0.0860	5	0.4301	25
Totals	0	0	1.9017	110

Otherwise, all computations to derive the regression parameters are identical to the linear regression described earlier. Again, we use index numbers, calculate the deviations to the mean values (e.g., log (p) and t), and compute the necessary products in column six and seven.

The regression line coefficients are then calculated as:

$$\log(\hat{\beta}) = \frac{\sum \log(p)\cdot t}{\sum t^2} = \frac{1.9017}{110} = 0.0173$$

$$\log(\hat{\alpha}) = \log(\overline{\text{Pop}}) - \log(\hat{\beta})\cdot \overline{T} = 4.8484 - 0.0173 \times 6 = 4.7447$$

Using the estimated regression coefficients, we can develop Boone County's geometric population model based on observed data for the years 1990 to 2000 as follows:

$$\log(\text{Pop}_n) = 4.7447 + 0.0173 \cdot T_n$$

Note that the estimated population model for Boone County is still in the logarithmic form. Although the model can be used to project Boone County's population for future years, it is important to recognize that these population projections will also be in their logarithmic form. Before getting meaningful population projections, we must convert the estimated population values back by taking the antilogarithm, the inverse operation of the logarithm. For instance, Boone County's population can be projected using the estimated regression model in logarithmic form for the years 2001, 2002, and 2003: [1]

$$\log(\widehat{\text{Pop}}_{2001}) = 4.7447 + 0.0173 \times 12 = 4.9522 \Rightarrow 10^{4.9522} = 89{,}571$$
$$\log(\widehat{\text{Pop}}_{2002}) = 4.7447 + 0.0173 \times 13 = 4.9695 \Rightarrow 10^{4.9695} = 93{,}208$$
$$\log(\widehat{\text{Pop}}_{2003}) = 4.7447 + 0.0173 \times 14 = 4.9867 \Rightarrow 10^{4.9867} = 96{,}994$$

Immediately you see that the results in their logarithmic form (e.g., 4.9522) have no direct meaning for planning purposes. Given that we used the base 10 logarithm, we get population projections by taking the inverse, or the antilogarithm, which is done as 10^{Pop_n}. [2]

Alternatively, we can write the geometric population model in its original form by taking the antilogarithms of the estimated regression line:

$$\text{Pop}_n = 55{,}553 \times 1.0406^{T_n}$$

The advantage of doing so is that we now have the population model in a form containing the constant annual population growth rate, r. For Boone County, the constant annual growth factor $(\hat{\beta})$ is 1.0406. The actual growth rate is then computed as (growth factor—1.0) and equals 0.0406 for Boone County. This indicates that between 1990 and 2000, the population in Boone County grew annually by a constant rate of 4.06%.

In a last step, we now project Boone County's population based on the geometric population model:

$$\widehat{\text{Pop}}_{2001} = 55{,}553 \times 1.0406^{12} = 89{,}571$$
$$\widehat{\text{Pop}}_{2002} = 55{,}553 \times 1.0406^{13} = 93{,}208$$
$$\widehat{\text{Pop}}_{2003} = 55{,}553 \times 1.0406^{14} = 96{,}994$$

Using either the geometric model or the transformed logarithmic version of the geometric model will result in identical population projections. Furthermore, the projections are slightly higher than the projections using the linear population model. The geometric model with its constant growth rate assumes faster growing populations in the later years.

---

① Please note that presented results have been calculated in a spreadsheet using more than 10 digits after the decimal points. Given that presented logarithm values indicate only 4 digits after the decimal point, some discrepancies to the final population projections will become apparent by recalculating the population projections using the rounded four digits after the decimal point values.

② The general form of the logarithm is: $y = \log_a(z) \Leftrightarrow a^y = z$, where a is called the base.

We conclude this section with the inclusion of the adjustment factor into the geometric population model for Boone County as:

$$\text{Pop}_n = 55{,}553 \times 1.0406^{T_n} + \text{ADJUST}$$

In similar fashion to the linear population model, the ***adjustment factor*** guarantees that the observed population for the last year data were available, i.e., year 2000 for Boone County, matches the estimated population for this particular year. The adjustment factor again for Boone County for 2000 is calculated as:

$$\text{ADJUST} = \text{observed Pop}_{2000} - \text{estimated Pop}_{2000} = 85{,}991 - 86{,}075 = -84$$

Based on the adjusted geometric population model, the projected population for Boone County for 2001, 2002, and 2003 is:

$$\text{Pop}_{2001} = 55{,}553 \times 1.0406^{12} - 84 = 89{,}487$$
$$\text{Pop}_{2002} = 55{,}553 \times 1.0406^{13} - 84 = 93{,}124$$
$$\text{Pop}_{2003} = 55{,}553 \times 1.0406^{14} - 84 = 96{,}910$$

3.4.5 Parabolic Population Model

The main assumption for the parabolic population model, like for the geometric model, is that under certain circumstances the population of an area is not expected to follow a linear growth path. The general equation for the parabolic curve is given in Eq. (3.21):

$$y = a + b_1 \cdot x + b_2 \cdot x^2 \tag{3.21}$$

The equation can be rewritten in Eq. (3.22) as a population model:

$$\text{Pop}_n = \alpha + \beta_1 \cdot (T_n) + \beta_2 \cdot (T_n^2) \tag{3.22}$$

where,

Pop_n — population in year n (dependent variable);

α — intercept;

β_1 and β_2 — coefficients of the parabolic curve;

T_n — index number for year n.

The more specific reason for choosing a parabolic over, for example a geometric population model, lies in the fact that the parabolic population model allows the incremental population growth (e.g., annual change in population expressed as people per year) to increase or decrease over time. (Remember that in a geometric model which assumes a constant growth rate over time, the annual population increase or decline expressed in people per year is always increasing).

This change in functional flexibility comes from the use of a linear and a nonlinear component in the parabolic model. Generally, the parabolic population curve is a quadratic function. As such, the signs (e.g., plus or minus) of estimated parameters determine if a population incrementally grows (declines) at increasing or decreasing rates. Given that we have two parameters to be estimated and each can have a positive or a negative sign, there are four different growth rate cases (Table 3.13).

Table 3.13 Effects of the signs of slope parameters on population growth/decline

Case	Sign of Linear Slope Parameter (β_1)	Sign of Nonlinear Slope Parameter (β_2)	Effects on Population Growth
Ⅰ	positive	positive	increasing incremental population growth concaves upward
Ⅱ	positive	negative	decreasing incremental population decline concaves downward
Ⅲ	negative	positive	decreasing incremental population growth concaves upward
Ⅳ	negative	negative	increasing incremental population decline concaves downward

The effects of the signs of the parameters β_1 and β_2 can easily be graphed by expanding the square term Time^2 by $(\beta_1 / 2\beta_2)$ and rewriting the parabolic population as:

$$\mathrm{Pop}_n = \beta_2 \left(T_n + \frac{\beta_1}{2\beta_2} \right)^2 + \alpha - \frac{\beta_1^2}{4\beta_2} \tag{3.23}$$

This is shown graphically below (Fig. 3.10). But note that only positive values for population and time are allowed.

Altogether, the parabolic model has three coefficients: the intercept with the y-axis, α, the coefficient for the linear term, β_1, and the coefficient for the nonlinear component, β_2. The parabolic curve can be estimated using ordinary least-square (OLS) regression techniques. However, adding a second variable (e.g., Time^2) to the right-hand side of the equation adds substantial computational complexity. Rather than doing the computation by hand as with the first two population models, we need to use a statistical software package. For example, we use SPSS software package to get the estimated regression coefficients for the

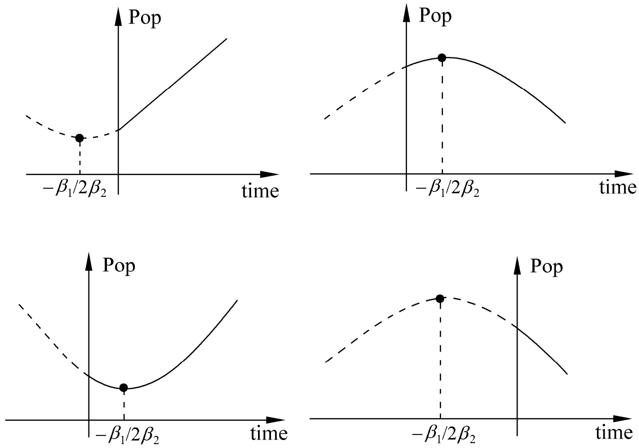

Figure 3.10 Effects of the signs of slope parameters

parabolic population model using the Boone County population data for 1990 to 2000. The three estimated parameters are:

$$\hat{\alpha} = 56,017$$
$$\hat{\beta}_1 = 1,962$$
$$\hat{\beta}_2 = 72$$

We now use these three estimated parameters and set up the Boone County parabolic population model:

$$\text{Pop}_n = 56,017 + 1,962 \cdot T_n + 72 \cdot T_n^2$$

Given that the slope parameters are both positive, the parabolic model for Boone County projects a population increase at an increasing rate. In particular, Boone County's population based on the parabolic model is projected as

$$\widehat{\text{Pop}}_{2001} = 56,017 + 1,962 \times 12 + 72 \times 144 = 89,929$$
$$\widehat{\text{Pop}}_{2002} = 56,017 + 1,962 \times 13 + 72 \times 169 = 93,691$$
$$\widehat{\text{Pop}}_{2003} = 56,017 + 1,962 \times 14 + 72 \times 196 = 97,598$$

The annual absolute increases in population growth are 3,762 and 3,906 between 2001 and 2002 and between 2002 and 2003, respectively.

We again have the choice of including an adjustment factor into the population model. As already demonstrated, the adjustment factor is calculated as the difference of the observed population for Boone County in 2000 and the projected population for the same year using the parabolic population model. Here, the adjustment factor is calculated as: $85,991 - 86,312 = -321$. In the next step, we add the adjustment factor to the parabolic population of Boone County.

$$\text{Pop}_n = 56,017 + 1,962 \cdot T_n + 72 \cdot T_n^2 - 321$$

And with the adjusted parabolic population model, the population projections for Boone County for 2001, 2002, and 2003 are calculated as:

$$\text{Pop}_{2001} = 56,017 + 1,962 \times 12 + 72 \times 144 - 321 = 89,609$$
$$\text{Pop}_{2002} = 56,017 + 1,962 \times 13 + 72 \times 169 - 321 = 93,371$$
$$\text{Pop}_{2003} = 56,017 + 1,962 \times 14 + 72 \times 196 - 321 = 97,276$$

The population models discussed so far have one thing in common. They all allow unlimited population growth or decline. In other words, there are no boundaries. Populations could grow indefinitely. Alternatively, unlimited decline would lead to the extinction of a population in a region. To avoid this fallacy, demographers apply contain upper and/or lower limits or boundaries to population models.

You can easily imagine that any region has limited carrying capacity, which is determined by the boundary of land area and other factors. The term carrying capacity, in this context, refers to the maximal population size that an area can support without reducing its ability to support the population in the future.[1] We will discuss the carrying capacity in further detail in Chapter 6. Setting an upper limit avoids projecting population growths that are beyond a region's carrying capacity.

Many towns, cities, or counties face the challenge to provide the necessary infrastructure (e.g., roads, water, sewer, and electricity among others), schools, libraries, housing, jobs, and recreational facilities for a growing population. On the other hand, places rarely die out completely and become ghost towns. Independent of socio-economic and political trends, people are attached to places where they grew up and spent their childhood. It is therefore implausible to anticipate that a population declining trend will lead to a population that will vanish over time.

The idea of setting upper ceilings and lower bounds to an area's population growth/decline is realized in several different population models: the logistic model, the modified exponential model, and the Gompertz model. However, in practice these models are rarely applied because setting ceilings is notoriously difficult to do. If the pasts do not provide reasonable upper and lower limits, setting ceilings is more often guessing than a methodological approach. In the following section we discuss one of these "constraint" population models—the

---

[1] Source: Population, Sustainability, and Earth's Carrying Capacity: A framework for estimating population sizes and lifestyles that could be sustained without undermining future generations, Gretchen C. Daily and Paul R. Ehrlich (1992), http://dieoff.org/page112.htm.

s-shaped logistic population model.

3.4.6 Logistic Population Model

The general form of the s-shaped logistic curve was first introduced by P. F. Verhurst, a Belgian mathematician in the 19th Century. Its popularity for population projections during the first part of the 20th century has been promoted by the work of Raymond Pearl and Lowell Reed (Klosterman, 1990). Although conceptually striking, the logistic model requires predetermining upper/lower population boundaries, which makes it less used than simpler models. Nevertheless and for populations with changing growth rates, the logistic population model still may deliver accurate population forecasts. This model may be of use, in particular, when an initial period of slow growth is followed by a period of rapid growth, which finally leads to a period of stagnating growth that levels off at an upper bound.[①]

Keyfitz (1968) gave the equation for a logistic curve as

$$Y = \frac{c}{1 + a \cdot \mathrm{e}^{-bX}} \tag{3.24a}$$

where,

X — the independent variable;

Y — the dependent variable;

a and b — parameters;

c — growth ceiling constant.

Setting $c \to c, a \to a \cdot c$, and $\mathrm{e}^{-b} \to b$, we can simplify the logistic curve as

$$\begin{cases} Y = \dfrac{1}{\dfrac{1}{c} + ab^X} \\ \dfrac{1}{Y} = \dfrac{1}{c} + ab^X \end{cases} \tag{3.24b}$$

We then get the logistic curve equation in a form that is familiar to us. It is important that although the predetermined growth ceiling is set at a parameter value, c, in the logistic curve function, the growth limit is given as its reciprocal value, $1/c$.

The population logistic model can thus be written as

① Smith et al., 2000: 170 – 171.

$$\frac{1}{\text{Pop}_n} = \frac{1}{c} + \alpha \cdot \beta^{T_n} \tag{3.25}$$

where,

Pop_n — Population in year n;

T_n — Index number for year n;

β — constant population growth factor;

α — parameter;

$1/c$ — reciprocal of the preset upper asymptotic population ceiling.

The logistic curve is applicable for scenarios with an upper growth limit of the population, as well as scenarios with a lower growth limit. The difference depends solely on the value of the β parameter. For $0 < \beta < 1$, we have the case of an upper growth limit, for $\beta > 1$ we have analogously, a lower growth limit. This is graphically shown in Fig. 3.11. We can rewrite the logistic population model Eq. (3.25) as

$$\text{Pop}_n = \frac{c}{1 + \alpha \cdot c \cdot \beta^{T_n}} \tag{3.26}$$

and for a large T_n:

when $0 < \beta < 1$, β^{T_n} approaches zero, Pop_n approaches c;

when $\beta > 1$, β^{T_n} approaches infinity, Pop_n approaches zero.

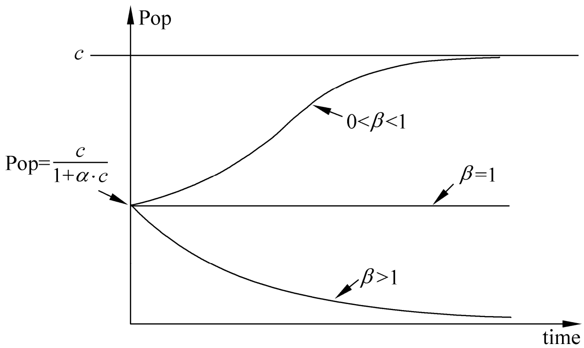

Figure 3.11 Effects of parameter β on population projections in the logistic model

We can transform the logistic curve to the linear form by taking the logarithms of Eq. (3.25):

$$\log\left(\frac{1}{\text{Pop}_n} - \frac{1}{c}\right) = \log(\alpha) + \log(\beta) \cdot T_n$$

where,

$\log(\beta)$ — slope of the population trend line in logarithmic form;

$\log(\alpha)$ — intercept of the trend line;

$\log(1/\mathrm{Pop}_n - 1/c)$ — log of the difference between the inverse of the population size and the inverse of the population ceiling.

The transformation of the logistic population model using logarithms allows us to apply the linear regression technique for population projection. Let us assume that Boone County's upper growth limit is 250,000 people,① the model solution is illustrated in Table 3.14.

For Boone County, the parameter estimates are:

$$\log(\hat{\beta}) = \frac{\sum \log\left(\frac{1}{p} - \frac{1}{c}\right) \cdot t}{\sum t^2} = \frac{-2.6599}{110} = -0.0242$$

$$\log(\hat{\alpha}) = \log\left(\frac{1}{\overline{\mathrm{Pop}}} - \frac{1}{c}\right) - \log(\hat{\beta}) \cdot \overline{T}$$

$$= -4.994 - 0.0242 \times 6 = -4.8492$$

The parameters $\hat{\alpha}$ and $\hat{\beta}$ then can be estimated by taking the antilogarithms of above parameter estimates.

$$\hat{\beta} = \text{anti}\log(\log\hat{\beta}) = \text{anti}\log(-0.0242) = 10^{-0.0242} = 0.9458$$

$$\hat{\alpha} = \text{anti}\log(\log\hat{\alpha}) = \text{anti}\log(-4.8492) = 10^{-4.8492} = 0.0000142$$

The final logistic population model for Boone County, KY, including the parameter estimates is

$$\frac{1}{\mathrm{Pop}_n} = \left(\frac{1}{250{,}000} + 0.0000142 \times 0.9458^{T_n}\right), \text{or}$$

$$\mathrm{Pop}_n = \frac{1}{\frac{1}{250{,}000} + 0.0000142 \times 0.9458^{T_n}}$$

Using this model, we can then project Boone County's population for the years 2001, 2002, and 2003 as follows:

① The upper growth limit is set at 250,000 people because Boone County is fast growing and because Boone County is partly rural with a large potential for future growth.

$$\frac{1}{\hat{\text{Pop}}_{2001}} = \frac{1}{250{,}000} + 0.0000142 \times 0.9458^{12} = 0.00001125 \Rightarrow \hat{\text{Pop}}_{2001} = 88{,}856$$

$$\frac{1}{\hat{\text{Pop}}_{2002}} = \frac{1}{250{,}000} + 0.0000142 \times 0.9458^{13} = 0.00001086 \Rightarrow \hat{\text{Pop}}_{2002} = 92{,}070$$

$$\frac{1}{\hat{\text{Pop}}_{2003}} = \frac{1}{250{,}000} + 0.0000142 \times 0.9458^{14} = 0.00001049 \Rightarrow \hat{\text{Pop}}_{2003} = 95{,}331$$

Table 3.14 Logistic population model calculations

	Observed Population (Pop)	Reciprocal Population Value (1/Pop)	Reciprocal Difference (1/Pop – 1/*c*)	Log of Difference (log(1/Pop – 1/*c*))	Index Numbers (*T*)
	57,589	0.00001736	0.00001336	– 4.874	1
	60,574	0.00001651	0.00001251	– 4.903	2
	62,897	0.00001590	0.00001190	– 4.924	3
	65,318	0.00001531	0.00001131	– 4.947	4
	67,554	0.00001480	0.00001080	– 4.966	5
	70,017	0.00001428	0.00001028	– 4.988	6
	72,860	0.00001372	0.00000972	– 5.012	7
	76,162	0.00001313	0.00000913	– 5.040	8
	79,818	0.00001253	0.00000853	– 5.069	9
	83,349	0.00001200	0.00000800	– 5.097	10
	85,991	0.00001163	0.00000763	– 5.118	11
Totals	782,129			– 54.938	66

Upper Population Limit (*c*):	250,000
Reciprocal Upper Limit (1/*c*):	0.000004
log (1/Pop – 1/*c*) Mean:	– 4.994
Mean Index Number:	6

Deviations from Mean Values		Necessary Cross-products	
from Population, log(1/*p* – 1/*c*)	from Index Number, *t*	log(1/*p* – 1/*c*) ·*t*	t^2
0.1203	– 5	– 0.6014	25
0.0915	– 4	– 0.3662	16
0.0698	– 3	– 0.2095	9
0.0478	– 2	– 0.0956	4
0.0279	– 1	– 0.0279	1

	0.0064	0	0.0000	0
	– 0.0178	1	– 0.0178	1
	– 0.0452	2	– 0.0904	4
	– 0.0748	3	– 0.2244	9
	– 0.1027	4	– 0.4108	16
	– 0.1232	5	– 0.6160	25
Totals	0	0	– 2.6599	110

Note that the projected population for Boone County based on the logistic model is calculated as its reciprocal value and needs to be converted back, which is shown on the right-hand side.

Similar to the previous models, we can adjust the logistic population model and recalculate the population projections for Boone County. Using the estimated logistic population model, the projected population for the year 2000 is 85,693. Thus, the logistic population model must be adjusted by 298 upwards. This is done in the easiest way by simply adding 298 to the outcome of the "unadjusted" logistic population model, or:

$$\hat{\text{P}}\text{op}_{2001} = 88,856 + 298 = 89,154$$
$$\hat{\text{P}}\text{op}_{2002} = 92,070 + 298 = 92,368$$
$$\hat{\text{P}}\text{op}_{2003} = 95,331 + 298 = 95,629$$

The same outcome is achieved by incorporating the adjustment factor into the logistic population model. In the logistic model where population is expressed in its reciprocal value, the adjustment term is calculated as:

$$\begin{aligned}\text{ADJUST} &= \frac{1}{\text{estimated Pop}_{2000}} - \frac{1}{\text{observed Pop}_{2000}} \\ &= \frac{1}{85,693} - \frac{1}{85,991} \\ &= 0.000000041\end{aligned}$$

Incorporating the adjustment factor into the logistic population model, the adjusted model becomes:

$$\frac{1}{\text{Pop}_n} = \frac{1}{250,000} + 0.0000142 \times 0.9458^{T_n} + 0.000000041$$

Also, note that the inclusion of the adjustment factor will move the upper (or lower) limit upwards (or downwards) by the value of the adjustment factor.

Over the past few pages, you have been introduced to six different population extrapolation methods. Two of them use simple ratios and four are based on more complex regression analysis. We further see that these four more complex methods use the “least-square criterion” to estimate the regression parameters. In Table 3.15 we compare population projections using all six different models for Boone County, KY for the years 2001 through 2010. If applicable, an adjustment factor is included (e.g., as shown in the four regression models). For a better comparison of each model’s functional forms and characteristics, we summarized some key concepts of these extrapolation models into Table 3.16.

As a first impression, all projected results seem to be reasonable considering the fast population growth of Boone County during the 1990’s. The share of growth

Table 3.15 Comparison of “adjusted” population projections for Boone County, KY

Year	Share of Growth	Shift-Share	Linear	Geometric	Parabolic	Logistic
2001	–	–	88,817	89,487	89,609	89,154
2002	–	–	91,643	93,124	93,371	92,368
2003	–	–	94,469	96,910	97,276	95,629
2004	–	–	97,295	100,849	101,326	98,934
2005	–	–	100,121	104,947	105,519	102,278
2006	–	–	102,947	109,213	109,857	105,656
2007	–	–	105,773	113,651	114,338	109,064
2008	–	–	108,599	118,270	118,963	112,497
2009	–	–	111,426	123,076	123,732	115,949
2010	112,617	117,851	114,252	128,078	128,645	119,416

method and the linear population model are at the low end of projections with 112,617 and 114,252 for 2010 respectively. The parabolic and geometric models have the highest ones with 128,645 and 128,078 for 2010. As mentioned earlier, models that rely on growth rates, such as the geometric and parabolic population models, have faster growing populations in later years which is clearly apparent in Table 3.15. But how can we determine which model provides the “best” results?

As a first step, you can visually examine observed population data to identify the growth pattern. Of course, having only few data points on a scatter plot makes it difficult to identify a pattern. Data for a longer time period could readily help identify if the visually observed pattern is of linear or geometric nature.

A more sophisticated input evaluation criterion uses the coefficient of relative variation (CRV). This is based on the idea of finding the curve that provides the closest match to observed historic data; this method compares the actual trend in observed historic data to the assumed trend for each extrapolation method. In other words, We compare the observed historic data to the estimated data derived from a trend curves. The CRV is defined as the ratio of the standard deviation to the mean:

$$\mathrm{CRV} = \frac{\text{standard deviation}}{\text{mean}} = \frac{s}{\bar{x}} \tag{3.27}$$

As a common measure of dispersion, it measures how dispersed our data are around a measure of central tendency, e.g., the mean. The closer the curve fits the historic population data, the less dispersed the data, which corresponds to a lower CRV. For Boone County, the CRV calculations are in Table 3.17. Please note that for these calculations only the historic data (for the index numbers 1 through 11) are included.

Following the criteria that the lowest CRV provides the best fit to observed population data, the logistic curve would be the best choice. However, our

Table 3.16 Summary characteristics of selected extrapolation methods

Model	Population Model	Regressand	Regressors	Estimated Parameters
Share of Growth	$\mathrm{Pop}_{m,\mathrm{ty}} = \mathrm{Pop}_{m,\mathrm{ly}} + \mathrm{growthshare}(\mathrm{Pop}_{n,\mathrm{ty}} - \mathrm{Pop}_{n,\mathrm{ly}})$	N/A	N/A	N/A
Shift-Share	$\mathrm{Pop}_{m,\mathrm{ty}} = \mathrm{Pop}_{n,\mathrm{ty}}\left[\mathrm{share}_{\mathrm{ly}} + \left(\frac{\mathrm{years}_{\mathrm{pp}}}{\mathrm{years}_{\mathrm{bp}}}\right)(\mathrm{share}_{\mathrm{ly}} - \mathrm{share}_{\mathrm{by}})\right]$	N/A	N/A	N/A
Linear	$\mathrm{Pop}_n = \alpha + \beta \cdot \mathrm{T}$	Pop_n	T_n	α, β
Geometric	$\mathrm{Pop}_n = \alpha \cdot \beta^{\mathrm{Time}}$	$\log \mathrm{Pop}_n$	T_n	$\log\alpha, \log\beta$
Parabolic	$\mathrm{Pop}_n = \alpha + \beta_1 \cdot \mathrm{Time} + \beta_2 \cdot \mathrm{Time}^2$	Pop_n	T_n and T_n^2	α, β_1, β_2
Logistic	$1/\mathrm{Pop}_n = 1/(c + \alpha \cdot \beta^{\mathrm{Time}})$	$\log(1/\mathrm{Pop}_n - 1/c)$	T_n	$\log\alpha, \log\beta$

Continued

Model	Characteristics
Share of Growth	small target area grows/declines according to a *predetermined and constant population growth share* of the smaller area in comparison to a larger region—requires population projection of larger comparison region
Shift-Share	small target area population is determined based on base year and launch year population shares of the smaller area to a larger region—requires population projection of larger comparison region
Linear	*constant absolute growth* increments $\beta: \mathrm{Pop}_{n+1} - \mathrm{Pop}_n$
	population increase: $+\beta$; population decline: $-\beta$
Geometric	*constant growth factor* $\beta = 1 + r$, where r is growth rate, e.g., $\mathrm{Pop}_{n+1} / \mathrm{Pop}_n = \text{constant} = 1 + r$
	“the rise of the increment in population is related to the size of the population and increasing”
	ratio between increment in population / total population = constant
Parabolic	*increasing/decreasing growth increments* depending on β_1 and β_2
	β_1 : constant linear growth component, e.g., $\mathrm{Pop}_{n+1} - \mathrm{Pop}_n = \text{constan}$
	β_2 : constant non-linear growth; constant in second differences
	depends on β_2 : if $\beta_2 < 0 \rightarrow$ concave downwards sloping
	if $\beta_2 = 0 \rightarrow$ linear
	if $\beta_2 > 0 \rightarrow$ concave upwards
Logistic	*constant ratio of increments for reciprocal*, $(1/\mathrm{Pop}_n)$, of population values
	constant factor: $(1/c - 1/\mathrm{Pop}_{n+1})/(1/c - 1/\mathrm{Pop}_n) = \text{constant} = 0.9458$
	s-shaped curve with upper growth limit of $1/c$ where upper asymptotic growth limit equals c
	growing populations: $\alpha > 0$ and $0 < \beta < 1$
	declining populations: $\alpha > 0$ and $\beta > 1$

Table 3.17 Evaluation of population projections[1]

	Observed Population data	Linear	Geometric	Parabolic	Logistic
pop. limit	–	–	–	–	250,000
Alpha (α)	–	54,146	55,553	56,017	0.00001415
Beta (β_1)	–	2,826	1.0406	1,962	0.9458
beta 2 (β_2)	–	–	–	72	–
1	57,589	56,972	57,809	58,052	57,525
2	60,574	59,798	60,157	60,230	60,028
3	62,897	62,624	62,600	62,552	62,604
4	65,318	65,451	65,142	65,019	65,253
5	67,554	68,277	67,788	67,629	67,973
6	70,017	71,103	70,541	70,383	70,764
7	72,860	73,929	73,405	73,281	73,622
8	76,162	76,755	76,386	76,323	76,547
9	79,818	79,581	79,488	79,509	79,536
10	83,349	82,407	82,716	82,839	82,585
11	85,991	85,233	86,075	86,312	85,693
Standard Deviation (σ)	9,403.81	9,372.99	9,371.43	9,396.67	9,355.21
Mean (μ)	71,102.64	71,102.64	71,100.73	71,102.64	71,102.71
CRV	–	13.1823	13.1805	13.2156	13.1573
MAPE	–	0.9238	0.4744	0.4703	0.5850

(1) Projections in Table 3.17 do not include adjustment factors.

calculations also show that all CRVs are extremely close to each other, indicating that in the case of Boone County, each of the four extrapolation methods listed in Table 3.17 would provide, at least for short-term projections, similar results. But be aware that with longer projection horizons the gap between the individual population projection models widens and the choice among different population models becomes more significant.

The most commonly used evaluation criterion is the mean absolute percentage error (MAPE). The MAPE is an output evaluation criterion and compares projected population values to the observed population statistics.① For Boone County, we would compare the projected values to the observed values for index numbers 1 through 11. The MAPE is the average value of the sum of absolute values of errors expressed in percentage terms and can be written as:

① The literature alludes numerous other measures of forecast errors: (1) mean error (ME), (2) mean absolute error (MAE), (3) mean percentage error (MPE), (4) root mean square error (RMSE), (5) Theil's U statistic, and (6) Theil's delta statistic.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} |\,\mathrm{PE}_i\,|, \tag{3.28}$$

and

$$\mathrm{PE} = \frac{\hat{y}_i - y_i}{y_i} \cdot 100$$

where,

y_i — observed population values;

$\hat{y}_i$ — forecasted population values;

n — total number of observations;

PE — percentage error.

As you can see, the MAPE is calculated by averaging the percentage difference between the calculated values and the original observations. The result is an indication of the accuracy of the model when applied to the initial data set. The more closely the calculated values are to the observed values, the smaller the MAPE, and therefore, the better the model.

Going back to Table 3.17, we can see that the computed mean absolute percentage errors range from 0.4703 to 0.9238 for the four models. Based on the output evaluation criterion, the parabolic curve indicates the lowest MAPE. All MAPEs can be considered to be very low and very close to each other, making the final model choice less straightforward. While we have applied the MAPE to historic data, e.g., 1990 to 2000, its drawback is that we cannot apply it to check future projections simply because we do not have future census data. For this reason, MAPEs are normally calculated for comparing projection values with census numbers once the latter becomes available. For instance, we could check 1997 Boone County population projections for the year 2000 with actual census 2000 data.

Based on the observed Boone County population data and using the visual, the input evaluation, and the output evaluation criteria, there is strong evidence that the more complex population models do not clearly out perform the simpler linear population model. Thus, using more sophisticated models is not necessarily a guarantor for better projections. For Boone County, the observed population values indicate an unmistakable growth trend. Even more, by visual observation, we already can conclude that all individual observations lie very close to a linear trend line. This particular circumstance is the reason that any of the six extrapolation methods will produce reasonable population projections. However, such a clear-cut case, as with Boone county, is not the rule of thumb. Usually, the computed coefficients of relative variation (CRV) and mean absolute percentage errors (MAPE) provide at least some decision guidelines for which model to choose. Nevertheless, both methods require estimated and observed population data and as such can only be done for time periods for which population data are available.

Remark on R^2 values

Using statistical software packages and having the computer do the curve fitting, as part of the output, you will usually get a R^2 value. As we have already discussed, the R^2 measures the amount of variation in the observed population values as explained by time. Therefore, the higher the R^2, the better the fit of your estimated straight regression line to observed data.

R^2 values are only appropriate for comparing different population projection models when the regressands, e.g., the population variable on the left-hand side of the linearly transformed population regression model, are identical.① Given that the population extrapolation models vary widely in their regressands, the R^2 does not provide the means for a comparison of the goodness-of-fit of different extrapolation models. In other words, an R^2 from a regression using absolute population as a dependent variable (Pop_n) cannot be compared to the R^2 of the geometric model which uses population in logarithmic form ($\log(\mathrm{Pop}_n)$).

3.5 Cohort-Component Method

A second main method for many state and local governments to project an area's population is the cohort-component method. The cohort-component method provides detailed demographic information on why and how the population changes.

The **first step** in the cohort-component method is to divide the population into age and sex cohorts. Further stratification depends primarily on needs and data availability and could be done according to race and ethnicity. More detailed subdivisions could follow, for instance, the racial and ethnic breakdown used in the 2000 Census.

In the **second step**, fertility, mortality, and migration rates, are applied to each individual cohort. For each cohort we will project how the population will change over a predetermined time period. Then we can answer questions like:

(1) How will the cohort of female of age 20 – 24 years change over the next five years?

(2) What is the projected change in the total male population for a 10-year time period?

(3) How is the area's population as a whole projected to change?

Before we get started, there are some more considerations that need to be taken into account. First, all age-groups must be uniform in that the years in the cohorts (n) are identical. Very often, cohort-component models divide the population into five-year age cohorts. This level of detail keeps data and

① Gujarati,1995: 171.

computational requirements within manageable limits while still providing sufficient details. Second, the number of years in the projection intervals (z) should relate to the number of years in the cohorts (n). For instance, using five-year age cohorts would logically suggest projecting for five-year periods (e.g., $n = z$). The advantage is that one specific age cohort, e.g., 25 – 29 years, would advance over a five-year projection period to the next age cohort, e.g., 30 – 34 years. This is shown in Fig. 3.12 below.

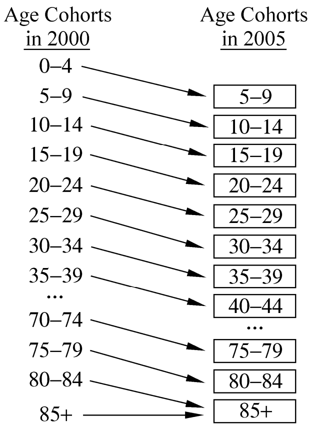

Figure 3.12 Age cohorts in the cohort-component model

And of course, all rates used in the cohort-component model must be adjusted to reflect five-year projection periods. For instance, the fertility rate for a particular female age-group, e.g., 491.7 for the age group of 20 – 24 years in Boone County, must reflect the appropriate time interval. Problems may arise when projecting five year age cohorts for let us say a three year time period as there is a clear mismatch of projection period and age cohort definition. However, using five-year age cohorts would also allow multiples of five-year projection intervals, for instance, 10 years, 15 years, etc.

Earlier in this chapter, we have referred to the individual components of change as births, deaths, and in- and out-migration. We further have discussed the individual rates that reflect these components namely fertility, survival, and migration rates. In this section, we now pull all required data for Boone County together and will develop a cohort-component model for the county. For Boone County, things are simplified, in that fertility, survival, and net migration rates by age and sex are available online at the Kentucky State Data Center & Kentucky Population Research (KSDC/KPR) at the University of Louisville Urban Studies Institute. The center provides these data for each of the 120 counties in Kentucky. The time interval (n) is five years. Particularly, we will project the 2000 Boone County population into the year 2005. The population is broken down into age-sex cohorts. The youngest five-year age group is 0 – 4 years, the oldest group lumps together all people over the age of 85. Furthermore, in the sample model we will be using net migration rates.

Table 3.18 contains the rates for the male population in Boone County and Table 3.19 contains the rates for the female population. It is noteworthy that fertility rates apply only for the female population of age 10 through 44, where the age is measured at the beginning of the five year interval.

Table 3.18 Male age-specific survival and migration rates per 1,000 persons

Beginning Age 2000	Ending Age 2005	Survival Rates	Net Migration Rates
Live Births	0 – 4	992.4	113.1
0 – 4	5 – 9	996.7	213.0
5 – 9	10 – 14	998.8	120.0
10 – 14	15 – 19	997.2	29.8
15 – 19	20 – 24	992.9	– 2.2
20 – 24	25 – 29	992.9	350.5
25 – 29	30 – 34	991.8	321.1
30 – 34	35 – 39	989.6	162.4
35 – 39	40 – 44	985.9	155.2
40 – 44	45 – 49	980.3	49.3
45 – 49	50 – 54	970.9	94.0
50 – 54	55 – 59	954.0	76.6
55 – 59	60 – 64	925.6	2.6
60 – 64	65 – 69	883.0	– 12.1
65 – 69	70 – 74	823.6	48.2
70 – 74	75 – 79	750.8	74.7
75 – 79	80 – 84	639.2	49.2
80 – 84	85 – 89	498.5	13.1
85 +	90 +	297.6	143.2

Table 3.19 Female age-specific fertility, survival, and migration rates per 1,000 persons

Beginning Age 2000	Ending Age 2005	Fertility Rates	Survival Rates	Net Migration Rates
Live Births	0 – 4	–	993.4	113.1
0 – 4	5 – 9	–	997.5	213.0
5 – 9	10 – 14	–	999.2	120.0
10 – 14	15 – 19	52.4	998.5	21.4
15 – 19	20 – 24	391.6	997.3	21.3
20 – 24	25 – 29	491.7	997.4	385.2
25 – 29	30 – 34	491.7	996.6	255.2
30 – 34	35 – 39	287.8	995.3	122.8

Continued

Beginning Age 2000	Ending Age 2005	Fertility Rates	Survival Rates	Net Migration Rates
35 – 39	40 – 44	86.4	993.2	160.6
40 – 44	45 – 49	9.5	989.1	10.2
45 – 49	50 – 54	–	983.5	142.4
50 – 54	55 – 59	–	973.0	67.5
55 – 59	60 – 64	–	957.0	18.0
60 – 64	65 – 69	–	932.0	– 12.1
65 – 69	70 – 74	–	897.0	48.2
70 – 74	75 – 79	–	850.4	74.7
75 – 79	80 – 84	–	765.3	49.2
80 – 84	85 – 89	–	637.3	13.1
85 +	90 +	–	381.4	143.2

Once the necessary population data and the age-specific fertility, survival, and net migration rates are collected and the model is conceptually prepared (e.g., deciding on the cohort breakdown and the time interval), we are ready to do all calculations in a spreadsheet. We will complete the model for the female population in Table 3.20. The cohort-component model is broken down into three parts:

(1) Columns one and two contain the initial female population in Boone County in 2000 $({}_n\mathrm{F}_x^{2000})$ broken down by age cohorts.

(2) Columns three to six include the age-specific survival, net migration, and fertility rates. We divided all initial rates from Tables 3.18 and 3.19 by 1000.

(3) Columns seven to thirteen contain the results from the calculations. These are:

① surviving female population in 2005 $({}_n\mathrm{SF}_{x+z}^{2005}$ — column seven),

② female deaths from 2000 to 2005 $({}_n\mathrm{DF}_x^{2000-2005}$ — column eight),

③ net migrating female population between 2000 to 2005 $({}_n\mathrm{NMF}_{x+z}^{2000-2005}$ — column nine),

④ female population in childbearing age $({}_n\mathrm{ARF}_x^{2005}$ — column ten),

⑤ number of projected births between 2000 to 2005 $({}_nB_x^{2000-2005}$ — column eleven),

⑥ projected female population in 2005 $({}_nF_{x+z}^{2005}$ — column twelve),

⑦ age cohorts in 2005.

Note that the first age cohort in future year 2005 is the 5 – 9 year cohort and the oldest age-group is 90 years and older. This is due to the fact that all female children of age 0 – 4 have moved after five years into the next higher age cohort, e.g., 5 – 9 years. The age cohort 0 – 4 in 2005 will be filled exclusively through births between 2000 and 2005.

Table 3.20 Female cohort-component module

Age in 2000	Female Pop. ${}_5F_x^{2000}$	Survival Rates ${}_5\mathrm{sr}_x$	Net Migration Rates ${}_5\mathrm{nmr}_x^{2000-2005}$	Fertility Rates ${}_5\mathrm{ASBR}_x^{2005}$	Adjusted Fertility Rates ${}_5\mathrm{abr}_x^{2005}$	Survive to 2005 ${}_5\mathrm{SF}_{x+z}^{2005}$	Deaths 2000 – 2005 ${}_5\mathrm{DF}_x^{2000-2005}$	Migrate 2000 – 2005 ${}_5\mathrm{NMF}_{x+z}^{2000-2005}$	At Risk Female Pop. ${}_5\mathrm{ARF}_x^{2005}$	Births 2000 – 2005 ${}_5B_x^{2000-2005}$	Projected Female Pop. ${}_5F_{x+z}^{2005}$	Age in 2005
1	2	3	4	5	6	7	8	9	10	11	12	13
0 – 4	3,347	0.9975	0.2130	–	–	3,339	8	713	–	–	4,052	5 – 9
5 – 9	3,458	0.9992	0.1200	–	–	3,455	3	415	–	–	3,870	10 – 14
10 – 14	3,309	0.9985	0.0214	0.0524	0.2220	3,304	5	71	3,377	750	3,375	15 – 19
15 – 19	2,940	0.9973	0.0213	0.3916	0.4417	2,932	8	63	2,999	1,324	2,995	20 – 24
20 – 24	2,490	0.9974	0.3852	0.4917	0.4917	2,484	6	959	3,446	1,694	3,443	25 – 29
25 – 29	3,100	0.9966	0.2552	0.4917	0.3898	3,089	11	791	3,886	1,515	3,881	30 – 34
30 – 34	3,621	0.9953	0.1228	0.2878	0.1871	3,604	17	445	4,057	759	4,049	35 – 39
35 – 39	4,003	0.9932	0.1606	0.0864	0.0480	3.976	27	643	4,632	222	4,619	40 – 44
40 – 44	3,879	0.9891	0.0102	0.0095	0.0048	3,837	42	40	3,897	19	3,876	45 – 49
45 – 49	3,188	0.9835	0.1424	–	–	3,135	53	454	–	–	3,589	50 – 54
50 – 54	2,730	0.9730	0.0675	–	–	2,656	74	184	–	–	2,841	55 – 59
55 – 59	1,913	0.9570	0.0180	–	–	1,831	82	34	–	–	1,865	60 – 64
60 – 64	1,419	0.9320	– 0.0121	–	–	1,323	96	– 17	–	–	1,305	65 – 69
65 – 69	1,182	0.8970	0.0482	–	–	1,060	122	57	–	–	1,117	70 – 74
70 – 74	1,110	0.8504	0.0747	–	–	944	166	83	–	–	1,027	75 – 79
75 – 79	784	0.7653	0.0492	–	–	600	184	39	–	–	639	80 – 84
80 – 84	527	0.6373	0.0131	–	–	336	191	7	–	–	343	85 – 89
85+	492	0.3814	0.1432	–	–	188	304	70	–	–	258	90 +
Total	43,492					42,092	1,400	5,050	26,295	6,283	47,142	

	Live Births 2000 – 2005 by Sex*	Survival Rates ${}_5\mathrm{sr}_x$	Projected Population by Sex	Child Deaths 2000 – 2005	Age in 2005
Female	3,065	0.9934	3,045	20	0 – 4
Male	3,218	0.9924	3,194	24	0 – 4
Total	6,283		6,238	45	
* Sex Ratio	1.05				

Total Projected Female Pop. F_{2005}
50,187

3.5.1 The Mortality Component

The first calculation we compute the female population likely to survive to the year 2005. Conceptually, the mortality component is presented in Fig. 3.13 below.

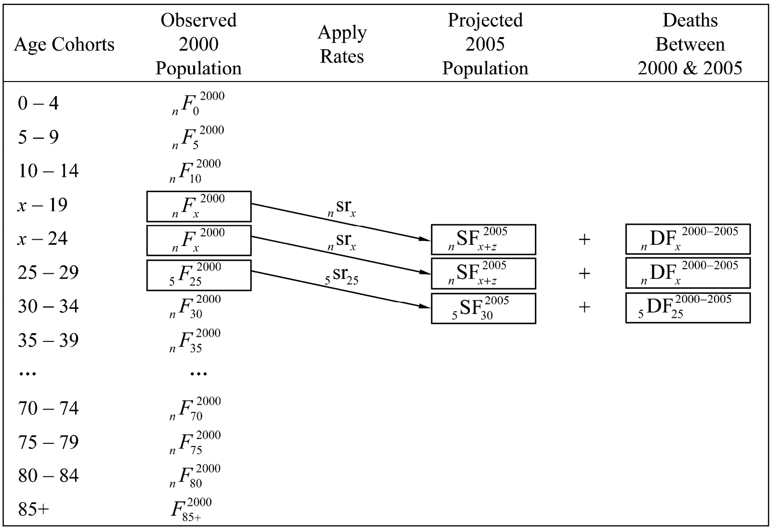

Figure 3.13 The mortality component of the cohort-component model

Depending on the age-specific survival rate $({}_n\mathrm{sr}_x)$, the female population from the initial year 2000 $({}_nF_x^{2000})$ either will move in the beginning of 2005 into the next age cohort $({}_n\mathrm{SF}_{x+z}^{2005})$ or will not survive from 2000 to 2005 $({}_n\mathrm{DF}_x^{2000-2005})$.

Computationally, this first step is done by multiplying the launch year female population in 2000 by its age-specific survival rate:

$$ {}_n\mathrm{SF}_{x+z}^{2005} = {}_nF_x^{2000} \cdot {}_n\mathrm{sr}_x , \tag{3.29} $$

where,

${}_nF_x^{2000}$ — female population in 2000;

${}_n\mathrm{SF}_{x+z}^{2005}$ — surviving female population in 2005;

${}_n\mathrm{sr}_x$ — age-specific survival rate, beginning age x, for five years age cohorts;

x — youngest age in a specific age cohort;

n — number of years in a specific age cohort (e.g., five years);

z — number of years in the projection interval (e.g., 2000 – 2005).

For the female age cohort 25 – 29 years in 2000, the surviving population in 2005 is

$$ {}_5\mathrm{SF}_{30}^{2005} = {}_5F_{25}^{2000} \cdot {}_5\mathrm{sr}_{25} = 3{,}100 \times 0.9966 = 3{,}089 $$

Taking the difference between the initial female population in 2000 and the surviving female population for that corresponding age cohort in 2005 will give us the number of female deaths in Boone County during the five-year projection period:

$$ {}_n\mathrm{DF}_x^{2000-2005} = {}_nF_x^{2000} - {}_n\mathrm{SF}_{x+z}^{2005} \tag{3.30} $$

where,

${}_n\mathrm{DF}_x^{2000-2005}$ — number of female deaths between 2000 and 2005, age cohort x.

For the female age cohort, $25-29$ in 2000, the number of females not surviving to the year 2005 would be projected as

$$ {}_5\mathrm{DF}_{25}^{2000-2005} = {}_5F_{25}^{2000} - {}_5\mathrm{SF}_{30}^{2005} = 3{,}100 - 3{,}089 = 11 $$

It should be emphasized that the youngest age cohort $0-4$ years in the projection year, 2005, is derived solely from cumulated births occurring between 2000 and 2005. This is described in detail in the fertility component section. Also, the oldest age cohort in 2005 now includes females aged 90 years or older. To be consistent with the 2000 age cohort definition, the two oldest age cohorts in 2005 (e.g., $85-89$ and $90+$) can be combined into one age cohort labeled $85+$. Alternatively, one could combine the two oldest age cohorts of the launch population (e.g., $80-84$ and $85+$ in 2000) into one cohort. The surviving population in the target year (e.g., $85+$ in 2005) is calculated by multiplying this combined population by the survival rate of the oldest population.

3.5.2 The Net Migration Component[①]

The second part of the calculations concentrates on deriving the net migrating female population for Boone County for the years 2000 to 2005. Calculating female net migrants versus calculating female in-migrants and female out-migrants has the advantage that it only requires one set of migration rates. However, by doing the net migration calculations, it is of importance to note whether the net migration rates refer to the initial launch year population or to the surviving target year population. Both approaches are possible and the choice is dependent upon how the net migration rates were derived, which is either by using the initial launch year population as the denominator or the surviving target year

① Conceptually, there is no difference between calculation net migration and in- and out-migration separately. In the example of Boone County, the choice between net migration and in- and out-migration calculations has been made dependent on data availability.

population as the denominator. In Boone County, the migration rates are per age-specific female cohort at the beginning of the five-year period.①

Another important and vital factor is the appropriate choice of the at-risk population. This addresses why people in- or out-migrate, which can depend upon socioeconomic factors internal or external to the area. First, it is theoretically justifiable that out-migration depends upon internal factors and therefore, also upon the area's population. Here, the population at risk, the population used to calculate the number of out-migrants, is the area's own population. The same logic does not hold for in-migration. The literature argues② that in-migration depends on factors external to the area of interest and, therefore, the appropriate choice of the population at risk to in-migrate should not be the area of interest. In the case of in-migration, the more appropriate population at risk is the population outside of the area under consideration. This can be, for example, the "adjusted U.S. population", which is derived by subtracting the area of interest's population from the U.S. population for a particular year. However, using, for example, the adjusted U.S. population as the population at risk to in-migrate to the area of interest explicitly implies that the in-migration rates must have been derived based on the adjusted U.S. population.

The choice of the appropriate population at risk is also necessary for calculating net migration rates. However, the choice can be different depending upon wether more people in-migrate than out-migrate or vice versa. Imagine a fast growing region with clearly far more people moving into this region than leaving it. In this situation, with in-migration being predominant, the net migration rates should be calculated using a population as base that lies outside the region, e.g., for example the adjusted U.S. population as previously described. Now picture a region which is losing population or growing at a very low rate. Here, net migration rates can be calculated based on the region's own population.

In practice, however, a far simpler approach is often used for calculating net migration rates. Net migration rates can easily be calculated as residuals by rearranging age-sex-specific demographic balancing equations. The only information necessary is the age-sex-specific population (P) for an area at two points in time (e.g., 1995 and 2000) and the number of deaths (D) and births (B) between these two points in time. We may then estimate the number of net migrants per age cohort x as:

$$ {}_n\mathrm{NM}_{x+z}^{1995-2000} = {}_nP_{x+z}^{2000} - {}_nP_x^{1995} - {}_nB_x^{1995-2000} + {}_nD_x^{1995-2000} \tag{3.31} $$

While in practice this is straightforward, we must keep in mind that conceptually net migration rates derived from residuals do not represent real probabilities, as is the case with fertility and survival rates.

① Source: http://ksdc.louisville.edu/kpr/pro/assumptions.htm.

② Smith et al., 2001:104 – 105.

The net migration rates for Boone County are listed under column four in Table 3.20. The Kentucky State Data Center calculated the rates using (1) the county's own population as the at-risk population and (2) the population at the beginning of the five-year time period. The calculations are again straight forward and are shown graphically in Fig. 3.14.

Age Cohorts	Observed 2000 Population	Apply Rates	Projected Female Migrants
0 – 4	${}_nF_0^{2000}$		
5 – 9	${}_nF_5^{2000}$		
10 – 14	${}_nF_{10}^{2000}$		
x – 19	$\boxed{{}_nF_x^{2000}}$	${}_n\text{nmr}_x^{2000-2005}$	
x – 24	$\boxed{{}_nF_x^{2000}}$	${}_n\text{nmr}_x^{2000-2005}$	$\boxed{{}_n\text{NMF}_{x+z}^{2000-2005}}$
25 – 29	$\boxed{{}_5F_{25}^{2000}}$	${}_5\text{nmr}_{25}^{2000-2005}$	$\boxed{{}_n\text{NMF}_{x+z}^{2000-2005}}$
30 – 34	${}_nF_{30}^{2000}$		$\boxed{{}_5\text{NMF}_{30}^{2000-2005}}$
35 – 39	${}_nF_{35}^{2000}$		
…	…		
70 – 74	${}_nF_{70}^{2000}$		
75 – 79	${}_nF_{75}^{2000}$		
80 – 84	${}_nF_{80}^{2000}$		
85+	F_{85+}^{2000}		

Figure 3.14 The net migration component of the cohort-component model

For Boone County, the age-specific number of female migrants is derived using the equation:

$${}_n\text{NMF}_{x+z}^{2000-2005} = {}_nF_x^{2000} \cdot {}_n\text{nmr}_x^{2000-2005} \tag{3.32}$$

where,

${}_n\text{NMF}_{x+z}^{2000-2005}$ — female population migrating between 2000 and 2005 per age cohort x;

${}_nF_x^{2000}$ — female population in 2000, age cohort x ("at-risk population");

${}_n\text{nmr}_x^{2000-2005}$ — net migration for age cohort x.

Again using the sample age cohort of females aged 25 – 29 in Boone County from Table 3.20, we calculated the number of female migrants for this age cohort as:

$${}_5\text{NMF}_{30}^{2000-2005} = {}_5F_{25}^{2000} \cdot {}_5\text{nmr}_{25}^{2000-2005} = 3{,}100 \times 0.2552 = 791$$

Of the females aged 30 – 34 in Boone County in 2005, 791 will be in-migrants.

3.5.3 The Fertility Component

The last portion of the cohort-component model calculates the number of births per age-specific female cohort. In particular, how many babies will be born to women of childbearing age, e.g., also referred to as women at risk. The outcome of these calculations will be used to project the number of females and males that go into the first age cohort 0 – 4 years in the target year 2005.

The fertility component of the model requires three individual steps which are

(1) to project the number of births per female age cohort;

(2) to aggregate all births and allocate this aggregated total between male and female births;

(3) to apply the survival rates to the cumulated male and female live births, projecting the number of males and females that will survive to the target year and form the youngest age cohort (e.g., 0 – 4 years in 2005).

The age-specific birth rates reported by the Kentucky State Data Center for Boone County are for a five-year period at the end of the five-year period. This requires adjusting the launch year female population in childbearing age by deaths and migration for the five-year period. The individual steps are graphically represented for the female cohort aged 25 – 29 in Fig. 3.15. The individual components are described following Fig. 3.15.

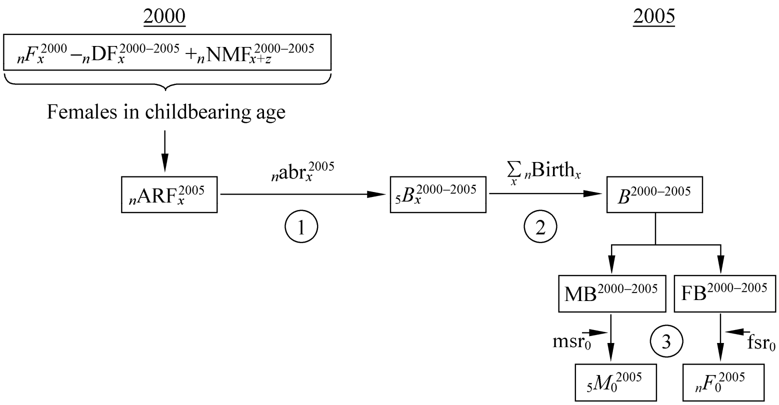

Figure 3.15 Fertility component of the cohort-component model

We see that the female population aged 25 – 29 is in a **first step** adjusted for deaths and migration and form the so-called at-risk female age cohort (ARF) aged 25 – 29. This at-risk female age cohort is then used to calculate the number of births per age cohort. Transforming this idea into equations we get:

$$ {}_n\mathrm{ARF}_x^{2005} = {}_nF_x^{2000} - (0.5 \cdot {}_n\mathrm{DF}_x^{2000-2005}) + {}_n\mathrm{NMF}_{x+z}^{2000-2005} \tag{3.33} $$

$$ {}_nB_x^{2000-2005} = {}_n\mathrm{ARF}_x^{2005} \cdot {}_n\mathrm{abr}_x^{2005} \tag{3.34} $$

$$ {}_n\mathrm{abr}_x^{2005} = \frac{({}_n\mathrm{ASBR}_x^{2005} + {}_n\mathrm{ASBR}_{x+5}^{2005})}{2} \tag{3.35} $$

where,

${}_n\mathrm{ARF}_x^{2005}$ — at-risk female population at the end of the projection interval, age cohort x;

${}_nB_x^{2000-2005}$ — births between 2000 and 2005, age cohort x;

${}_n\mathrm{abr}_x^{2005}$ — adjusted birth rate, age cohort x;

${}_n\mathrm{ASBR}_x^{2005}$ — age-specific birth rate, age cohort x.

Equation (3.33) determines the at-risk female population by adjusting the female launch year population (${}_nF_x^{2000}$) for deaths (${}_n\mathrm{DF}_x^{2000-2005}$) and female net migration (${}_n\mathrm{NMF}_{x+z}^{2000-2005}$). The adjustment of the deaths per female cohort by the factor 1/2 needs some more explanation. The assumption is that on average, a woman of childbearing age will stay half of the projection interval (e.g., $n/2$) in one age cohort and half in the next higher age cohort. For a five-year interval, a woman age 27 will stay three more years in the 25 – 29 cohort and the two remaining years in the 30 – 34 cohort. The direct result is that not all women belonging to the 25 – 29 cohort at the beginning of the interval will die at age 25 – 29 which requires adjusting the number of female deaths per cohort.

Equation (3.34) then calculates the projected births per age cohort by multiplying the at-risk female population (${}_n\mathrm{ARF}_x^{2005}$) by the corresponding adjusted birth rate (${}_n\mathrm{abr}_x^{2005}$). ① In section 3.3, we have discussed the age-specific birth rate (${}_n\mathrm{ASBR}_x^{\mathrm{year}}$). The concept for adjusting the age-specific birth rates is the same as for deaths. Women, on average, will only stay half of the projection interval (e.g., $n/2$) before moving into the next higher age cohort. For instance, the average age for women aged 25 – 29 years is 27.5 years, the middle year of this specific age cohort. On average, a woman will therefore stay 2.5 years in the age cohort 25 – 29 years before she advances into the age cohort 30 – 34 years, meaning that women spend half their time in one age cohort and the other half in the next higher age cohort.

① Given that the adjusted fertility rate is derived from age-specific birth rates, we alternatively also refer to it as the adjusted birth rate (${}_n\mathrm{abr}_x^{2005}$).

For Boone County females aged 25 – 29 in 2000, the births are computed as (see also column 10 and 11, Table 3.20):

$$
\begin{aligned}
{}_5\mathrm{ARF}_{25}^{2005} &= {}_5F_{25}^{2000} - (0.5 \times {}_5\mathrm{DF}_{25}^{2000-2005}) + {}_5\mathrm{NMF}_{30}^{2000-2005} \\
&= 3{,}100 - (0.5 \times 11) + 791 = 3{,}886
\end{aligned}
$$

$$
{}_5B_{25}^{2000-2005} = {}_5\mathrm{ARF}_{25}^{2005} \cdot {}_5\mathrm{abr}_{25}^{2005} = 3{,}886 \times 0.3898 = 1{,}515
$$

$$
{}_5\mathrm{abr}_{25}^{2005} = \frac{{}_5\mathrm{ASBR}_{25}^{2005} + {}_5\mathrm{ASBR}_{30}^{2005}}{2} = \frac{0.4917 + 0.2878}{2} = 0.3898
$$

The **second step** combines all births per age cohort into one aggregated figure, total births in Boone County between 2000 and 2005 ($B^{2000-2005}$). Having a cumulative figure for births in Boone County, we split the total births based on the historic male/female sex ratio at birth into cumulative male and female live births ($\mathrm{MB}^{2000-2005}$ and $\mathrm{FB}^{2000-2005}$).

This is done as follows:

$$
B^{2000-2005} = \sum_x {}_nB_x^{2000-2005} \tag{3.36}
$$

$$
\mathrm{MB}^{2000-2005} = \left[\frac{1.05}{1+1.05}\right] B^{2000-2005} \tag{3.37}
$$

$$
\mathrm{FB}^{2000-2005} = \left[\frac{1}{1+1.05}\right] B^{2000-2005} \tag{3.38}
$$

where,

$B^{2000-2005}$ — cumulative births between 2000 and 2005;

$\mathrm{MB}^{2000-2005}$ — cumulative male births between 2000 and 2005;

$\mathrm{FB}^{2000-2005}$ — cumulative female births between 2000 and 2005;

1.05 — historic male/female sex ratio at birth[①].

The idea here is to derive the total number of male and female babies born separately to women in Boone county of childbearing age between 2000 and 2005. In return, these two totals, e.g., $\mathrm{MB}^{2000-2005}$ and $\mathrm{FB}^{2000-2005}$, will be adjusted for infant mortality and finally, be used to build the youngest age cohort aged 0 – 4, in the target year 2005. Before we go to the last step, let us do the calculations of female and male births in Boone County. Beginning with step two, the actual calculations are added underneath Table 3.20.

---

① Source: http://www.odci.gov/cia/publications/factbook/print/us.html.

$$B^{2000-2005} = \sum_x {}_nB_x^{2000-2005} = 750 + 1,324 + 1,694 + 1,515 + 759 + 222 + 19$$
$$= 6,283$$

$$\text{MB}^{2000-2005} = 6,283 \times \left[\frac{1.05}{1+1.05}\right] = 6,283 \times 0.5122 = 3,218$$

$$\text{FB}^{2000-2005} = 6,283 \times \left[\frac{1}{1+1.05}\right] = 6,283 \times 0.4878 = 3,065$$

The calculations project a total of 3,218 male and 3,065 female births between 2000 and 2005 to all women of childbearing age in Boone County.

The fact that not all newborns will survive to the target year 2005 is shown in **step three**, where we adjust the number of male and female live births for infant mortality. This is done analogously to all other age cohorts, in that the cumulative live births are multiplied by a sex-specific survival rate.

$${}_nF_0^{\,2005} = \text{FB}^{\,2000-2005} \cdot {}_n\text{sr}_0 \tag{3.39}$$

where,

${}_nF_0^{2005}$ — youngest female age cohort aged 0 – 4 in 2005;

${}_n\text{sr}_0$ — survival rate for the age cohort 0 – 4.

We then compute the females in Boone County aged 0 – 4 in 2005 as:

$${}_5F_0^{2005} = 3,065 \times 0.9934 = 3,045$$

Boone County is projected to have 3,045 females in the youngest age cohort aged 0 – 4 in 2005.

3.5.4 Bringing All Components Together

So far, we have calculated the surviving female population, the net migrating female population, and the female births surviving to the year 2005. Now we are ready to bring all these individual pieces together in one equation in order to project the age-specific female population in 2005. But note that this final equation is not applicable for the youngest age cohort aged 0 – 4 in 2005 which comes exclusively form the fertility component:

$${}_nF_{x+z}^{\,2005} = {}_n\text{SF}_{x+z}^{\,2005} + {}_n\text{NMF}_{x+z}^{\,2000-2005} \tag{3.40}$$

Column twelve in Table 3.20 shows these final calculations. For Boone County, the projected female population aged 30 – 34 years in 2005 is:

$${}_5F_{30}^{2005} = {}_5\mathrm{SF}_{30}^{2005} + {}_5\mathrm{NMF}_{30}^{2000-2005} = 3,089 + 791 = 3,881$$ ①

The sum of all these projected age-specific cohorts in column twelve, Table 3.20, plus the result form the youngest age cohort aged $0-4$ in 2005 from the fertility component will give us the final result, the cumulative projected female population for Boone County in the target year 2005:

$$F^{2005} = \sum_x {}_nF_{x+z}^{2005} + {}_5F_0^{2005} = 47,142 + 3,045 = 50,187 \tag{3.41}$$

where,

F^{2005} — cumulative female population in 2005;

x — youngest age in a specific age cohort in 2005, e.g., $0-4$, $5-9$, $10-14$, etc.;

z — number of years in the projection interval (e.g., $2000-2005$).

The projected female population in Boone County for 2005 totals 50,187.

All calculations above refer to the female part of the cohort-component model. To get a complete small area model for Boone County, the same calculations need to be repeated for the county's male population. They are identical to what has been described for the female population. They are even simplified in that the male calculations do not include the fertility component. To show a complete cohort-component for Boone County, we added age-specific calculations for the male population in Table 3.21 without further elaborations. Adding total projected female and male population from the two cohort-component models will then give us Boone County's total projected population for the year 2005. It is calculated as:

$$P^{2005} = F^{2005} + M^{2005} = 50,187 + 49,266 = 99,453 \tag{3.42}$$

Hamilton and Perry (1962) proposed a short version of the cohort-component method, which apply **cohort-change ratios** (CCR) to the beginning population. These cohort-change ratios are usually calculated from the last two censuses. Given that censuses in the United States are ten years apart, the Hamilton-Perry method often projects five-year age groups in ten year intervals. Mortality and migration are combined into a single rate rather than treating them separately. Further simplification is possible by using child-woman ration instead of the age-specific birth rates in the fertility component. While the Hamilton-Perry method may be given preference where data are not readily available to build a more complex cohort-component model, a potential source of error is the use of constant growth rates—which in the case of particularly fast growing regions, can overestimate future populations.

---

① The difference of one female in this equation is due to rounding in the table.

Table 3.21 Male cohort-component module

Age in 2000	Male Pop. $_5M_x^{2000}$	Survival Rates $_5\mathrm{sr}_x$	Net Migration Rates $_5\mathrm{nmr}_x^{2000-2005}$	Survive to 2005 $_5\mathrm{SM}_{x+z}^{2005}$	Deaths 00 to 05 $_5\mathrm{DM}_x^{2000-2005}$	Migrate 00 to 05 $_5\mathrm{NMM}_{x+z}^{2000-2005}$	Projected Male Pop. $_5\mathrm{M}_{x+z}^{2005}$	Age in 2005
1	2	3	4	5	6	7	8	9
0 – 4	3,502	0.9967	0.2130	3,490	12	746	4,236	5 – 9
5 – 9	3,685	0.9988	0.1200	3,681	4	442	4,123	10 – 14
10 – 14	3,477	0.9972	0.0298	3,467	10	104	3,571	15 – 19
15 – 19	3,142	0.9929	– 0.0022	3,120	22	– 7	3,113	20 – 24
20 – 24	2,591	0.9929	0.3505	2,573	18	908	3,481	25 – 29
25 – 29	3,003	0.9918	0.3211	2,978	25	964	3,943	30 – 34
30 – 34	3,584	0.9896	0.1624	3,547	37	582	4,129	35 – 39
35 – 39	3,852	0.9859	0.1552	3,798	54	598	4,396	40 – 44
40 – 44	3.749	0.9803	0.0493	3,675	74	185	3,860	45 – 49
45 – 49	3,153	0.9709	0.0940	3,061	92	296	3,358	50 – 54
50 – 54	2,610	0.9540	0.0766	2,490	120	200	2,690	55 – 59
55 – 59	1,940	0.9256	0.0026	1,796	144	5	1,801	60 – 64
60 – 64	1,365	0.8830	– 0.0121	1,205	160	– 17	1,189	65 – 69
65 – 69	1,047	0.8236	0.0482	862	185	50	913	70 – 74
70 – 74	839	0.7508	0.0747	630	209	63	693	75 – 79
75 – 79	560	0.6392	0.0492	358	202	28	386	80 – 84
80 – 84	246	0.4985	0.0131	123	123	3	126	85 – 89
85 +	154	0.2976	0.1432	46	108	22	68	90 +
Total	42,499			40,899	1,600	5,173	46,072	

Cummulative Live Births 2000 – 2005 MB_{00-05}	Survival Rates $_5\mathrm{sr}_{2000}$	Projected Male Pop. M_{2005}	Child Deaths 00 to 05	Age in 2005
3,218	0.9924	3,194	24	0 – 4

Total Projected Male Pop. M_{2005}
49,266

3.6 Concluding Remarks

The population models have one thing in common. They assume that observed population trends can be carried over into the near future. As we have seen, this is repeatedly done for the population trend extrapolation methods by

(1) Gathering population data for the past years,

(2) Plotting these observed population data onto a scatter plot and inspecting them visually,

(3) Extrapolating observed trends into the near future, either using simple ratio methods as the share of growth or shift-share method or the more complex regression models such as the linear, geometric, parabolic, or logistic population model.

The strength of all trend extrapolation methods undoubtedly lie in their small data requirement. In most cases, total population figures for a number of past years will be sufficient to obtain an area's population projections for future years. This low data requirement makes extrapolation models very attractive for small areas where historical population data are not always readily available at a more detailed level.

Regarding the required modeling skills, trend extrapolation models vary significantly. While share of growth and shift-share models are conceptually and computationally easy to understand and implement, regression models such as the logistic model, involve higher mathematical skills. Nevertheless, understanding the linear transformation of the more sophisticated extrapolation models makes them easy to apply as the linear model.

On the other hand, extrapolation models have severe drawbacks. Given their use of limited and highly aggregated data, they lack any information about the different components of projected population growth/decline. In other words, we get no explanation on theoretical grounds for these projected population changes. The attractiveness of the very low data requirement, therefore, must be acknowledged as an intrinsic limitation to the projected results. Further, the use of highly aggregated data for past years treats factors like the area's economic, housing, and/or recreational attractiveness as external to the method.

Another difficult task is the choice of the appropriate timeframe for selecting past population statistics. If available, should we use populations for the last 10, 20, or even 50 years? This decision is less problematic for areas where data show a slow but steady growth pattern. The choice is more challenging for fast-growing areas or for areas with alternating population growth and decline patterns. What if the area grew generally at a slow rate for the last 50 years, but suddenly seven years ago began to decline? While typically the heaviest reliance is on the more recent data, there is no guarantee that the population will continue to decline. Other data might be of particular help for choosing the appropriate period of past years. For instance, improvements in the area's economic environment might have already indicated a turn in population growth. If this is the case, choosing only the last seven years would erroneously lead to further projected population decline while we know from outside sources that the population may more likely start growing again.

Another shortcoming of trend extrapolation methods is that they do not allow the ability to play out different future scenarios. For instance, how will population projections change if economic conditions, birth rates, migration patterns, and other factors change? We want to reemphasize that extrapolation models rely upon the assumption that observed past conditions are assumed to continue in the future. This assumption may or may not hold. As a direct consequence, the further into the future we project, the less reliable the extrapolation models become. The likelihood of a continuation of observed past population trends is greater when projecting only a few years into the future as when projecting population trends for as many as 20 – 30 years into the future. Therefore, population projections that go too far into the future must be read with reservation.

We want to conclude this section on extrapolation methods with the remark that in many cases, practitioners often use the simple and straight forward population models. Also in favor of extrapolation models is the fact that there is virtually no evidence that more complex methods outperform extrapolation models.

Population analysis is a challenging task and by far more than just running some population models. As Rayer puts it: [①]

"*The real challenge in population analysis is describing and projecting populations within 'reasonable' limits. We spend 90% of our time making sure the data make sense and if they do not, we make adjustments.*"

Computationally, the cohort-component method is easy and straightforward. Once the required data and rates are collected, its computations can be done using a spreadsheet. Cohort-component methods are widely used at all government levels. Its popularity mainly goes back to the amount of detail provided by the model. The breakdown of the target population into cohorts allows zooming in on specific parts of the population, e.g., females of age 30 to 34. For each of these cohorts, the components of the model, e.g., births, deaths, and migration, explain the reasons for population changes over time.

For many areas, the question is not only by how much the total population is projected to change over the next five years. For better planning, many local governments want to understand:

(1) Why is the population changing? Are the reasons for the expected changes mainly driven by births, deaths, and/or in- and out migration?

(2) How is the population changing? Is the population aging? Is the racial composition of the population changing?

While many of these questions can be answered directly from the

① Quote from personal communication with Dr. Stefan Rayer.

cohort-component model, it also means that with more detailed models the data requirements increase significantly. For a detailed population projection, population data by sex, age, and race/ethnicity as well as all fertility, survival, and in- and out- migration rates must be available. The computational requirements, time, and costs increase with the level of detail.

For planning purposes, many planning agencies will not get involved in the process of collecting and verifying data and computing the individual birth, survival, and migration rates. As planners, we prefer using readily available population data and rates to set up a cohort-component model. Using readily available data and rates provided by various governmental agencies, we are incorporating all assumptions into the cohort-component model. For example, the Boone County net migration rates were derived by using the area's own population as the population base. As such, we cannot use the adjusted U.S. population as the population at risk of migrating. Another example would be if the individual birth, survival, and migration rates have been calculated using the base population at the beginning, in the middle, or at the end of a specific time interval. For instance, had the survival rate been constructed using the base population at the beginning of a time interval, all following calculations must use the target population at the beginning of this particular time interval for consistency. For planners using pre-calculated rates can mean a higher degree of dependability on the assumptions made by the data collecting and rate calculating agency.

A last but critical point is the fact that all individual rates are calculated using historic data. Assuming that the observed trends in the components of growth and the demographic composition of the population remain constant for future time periods, we use rates computed from historic data and apply them to project future population growth. For instance, the computed birth rates based on the number of live births for the last period data available are used for projecting births for the next time period. The main assumption here is similar to that from the population trend models: past population trends can be carried on into the near future.

Each population projection method discussed in this chapter is applicable under certain situations. The choice of appropriate method should be a combination of purpose, time-money constraints, level of detail, and data availability. The most important of all is to ensure that input data are correct and reasonable. In reality, researchers spend most their time working on the data rather than running the models. Keep in mind that for population projections, you will spend hours and hours making sense out of your input data and in many cases, changes to the collected data are necessary before they can be used for population modeling purposes.

In general, there is virtually no evidence that more complex population methods, such as cohort-component and structural models, provide better population projections. Each method has strengths and weaknesses and each is based on a set of assumptions, which have an impact on the results. What more complex models do offer, however, is the ability to play out different future scenarios by using altering migration, birth, or survival rates. In addition, all these different rates applied in the cohort-component method could be trended themselves. For instance, observing migration rates on an annual basis for a longer period of time would allow to project migrations rates for future time periods.

Given that it is almost impossible to tell which of the described population projection methods would achieve “more accurate” projections under given conditions, it is very common to apply a mixture of different methods. Using averages derived from a mixture of methods is a more conservative way of projecting future population. The fact is that long-term trends are likely to regress towards the mean. In addition, it is common to provide an interval of projected populations rather than offering one exact population projection (e.g., point estimator). Calculating a series of low, middle, and high population projections allows to project populations within a range of values. Usually the middle series reflects what you believe is the most likely occurring population trend.

Projection errors decrease with population size. This is why many agencies use a stepwise approach of projecting populations. In a stepwise approach, state totals are calculated first. In a second step, county totals are calculated and the sum of all county totals must equal the state total. If not, adjustments to the county totals are made until their total equals the state total. More detailed calculations at the county-level, for instance, individual age cohorts are included in the third step. These more detailed projections are controlled by the county total.

Review Questions

1. What is the difference between population projections, forecasts, and estimates? From the U.S. Census Bureau website, are the inter-decennial population figures for the years 1991 – 1999 projected, forecasted, or estimated? Describe why?

2. Choosing the most appropriate projection method can depend on a variety of factors. Name at least five factors you think should be considered when choosing a projection method.

3. Briefly describe the four fundamental concepts of demographic analysis.

4. According to the demographic balancing equation, there are three components of change. Name these three components of change and explain how these components are being accounted for in the cohort-component model.

5. Trend extrapolation methods are very popular to project populations. Explain the rational behind all these extrapolation models. Explain the main

conceptual difference between the group of trend extrapolation models and the cohort-component method.

6. Under what circumstances would you consider the geometric population model as being appropriate to project population growth or decline? What does the slope coefficient of the geometric population model express? And what is the most obvious and important difference between the geometric population model and the logistic population model?

7. For each of the four trend extrapolation models discussed in chapter 3, we provided an adjusted form of the model. Explain the rational behind the inclusion of the adjustment factor into the model.

8. Describe in detail the net migration component of the cohort-component method. For a region that is losing population, what would you choose as denominator for calculating the net migration rate: the region's population or the population outside the region? Explain why.

9. What are the strengths and weaknesses of the cohort-component method?

10. Under what circumstances would you prefer an extrapolation model over a cohort-component model?

Exercises

You are hired as a planner for a small urban county, Sunshine County, and one of your first tasks is to update the county's demographic profile and to provide the county government with "reasonable" population projections until the year 2010.

1. Your first analysis is the graphical presentation of the county's population using a population pyramid (for detailed instructions on how to build a population pyramid, see Chapter 8). The county population data are listed in Table 3.22. What detailed information on Sunshine County does the population pyramid exhibit?

2. Calculating average annual absolute change (AAAC) and average annual percent change is a quick way of examining past observed population trends. Using the population data for the last twenty years, calculate the AAAC and AAPC for Sunshine County and interpret your results. In addition, project the county's total population for the year 2010 using the AAAC and AAPC.

3. A more sophisticated way of projecting populations uses trend extrapolation models. Using the data from Table 3.23, estimate the linear and geometric population model for Sunshine County. Compare your results with the AAAC and AAPC from above and project the county's total population for the year 2010 using both the linear and geometric extrapolation models.

4. Another quick way of projecting an area's future total population applies ratio methods, such as the share of growth and the shift-share method. Project the 2010 Sunshine County population using both the share of growth and the shift-share method, based on the information provided in Table 3.24.

Table 3.22 Sunshine County population by sex and age, 2000

Sunshine County, 2000	Male	Female
Total Population	19,185	19,856
Under 5 years	1,512	1,364
5 to 9 years	1,509	1,437
10 to 14 years	1,476	1,357
15 to 19 years	1,492	1,490
20 to 24 years	1,543	1,665
25 to 29 years	1,450	1,504
30 to 34 years	1,410	1,475
35 to 39 years	1,507	1,600
40 to 44 years	1,525	1,660
45 to 49 years	1,399	1,466
50 to 54 years	1,184	1,238
55 to 59 years	840	860
60 to 64 years	649	712
65 to 69 years	542	576
70 to 74 years	432	482
75 to 79 years	346	391
80 to 84 years	222	298
85 years and over	147	281

Table 3.23 Annual total population data for Sunshine County, 1980 – 2000

Year	Total Population	Year	Total Population
1980	26,065	1991	31,531
1981	26,611	1992	32,475
1982	26,759	1993	33,356
1983	27,283	1994	34,125
1984	27,794	1995	34,947
1985	27,917	1996	36,017
1986	28,453	1997	36,967
1987	28,976	1998	37,620
1988	29,682	1999	38,419
1989	29,992	2000	39,041
1990	30,508		

Table 3.24 Comparison of Sunshine County population to a benchmark region

Year	Benchmark Region	Sunshine County
1990	360,000	30,508
2000	400,000	39,041
2010	425,000	

5. The cohort-component model requires more detailed data than previous methods. Using Tables 3.22 and 3.25, which provide data on population by sex and age, birth rates, survival rates, and migration rates, set up the female cohort-component module and project the female age-specific population for the year 2010.

Table 3.25 Survival, birth, and net migration rates for Sunshine County, 2000 – 2005

Beginning Age in 2000	Ending Age in 2005	Survival Rates(2)		Net Migration Rates(3)		Birth Rates(4)
		Male	Female	Male	Female	
Live births(1)	0 – 4	992.4	993.4	26.7	6.7	–
0 – 4	5 – 9	996.7	997.5	147.3	127.3	–
5 – 9	10 – 14	998.8	999.2	96.7	76.7	–
10 – 14	15 – 19	997.2	998.5	73.8	53.8	64.9
15 – 19	20 – 24	992.9	997.3	135.2	115.2	363.4
20 – 24	25 – 29	992.9	997.4	112.4	92.4	551.1
25 – 29	30 – 34	991.8	996.6	– 14.5	– 14.5	488.3
30 – 34	35 – 39	989.6	995.3	68.9	48.9	240.6
35 – 39	40 – 44	985.9	993.2	98.9	78.9	75.2
40 – 44	45 – 49	980.3	989.1	23.3	3.3	7.2
45 – 49	50 – 54	970.9	983.5	86.6	66.6	–
50 – 54	55 – 59	954.0	973.0	7.4	– 12.6	–
55 – 59	60 – 64	925.6	957.0	61.4	41.4	–
60 – 64	65 – 69	883.0	932.0	52.0	32.0	–
65 – 69	70 – 74	823.6	897.0	48.4	28.4	–
70 – 74	75 – 79	750.8	850.4	123.2	103.2	–
75 – 79	80 – 84	639.2	765.3	117.3	97.3	–
80 – 84	85 – 89	498.5	637.3	214.1	194.1	–
85 +	90 +	297.6	381.4	69.7	49.7	–

(1) Cumulative live births during the 5-year period.
(2) Total survivors (those who do not die) per 1,000 persons over a 5-year period.
(3) Rates are per 1,000 persons at the beginning of the 5-year period.
(4) Total live births per 1,000 females over a 5-year period.

References

Armstrong, J.S. 1985. *Long-range Forecasting: From Crystall Ball to Computer*. 2nd. ed. New York, NY: John Wiley.

Batie, Sandra. 1989. Sustainable development: challenges to the profession of agricultural economics. *American Journal of Agricultural Economics*, 1,083 – 1,101.

Boone County, Kentucky. 2004. Official Website. Available online from http://www.boonecountyky.org/. Accessed September 2004.

Bowley, A. L. 1924. Birth and population in Great Britain. *The Economic Journal,* 34: 188 – 192.

Cannan, E. 1895. The probability of a cessation of the growth of population in England and Wales during the next century. *The Economist Journal*, 5: 506 – 515.

Central Intelligence Agency (CIA). 2004. The World Factbook: United States. Internet; Available from http://www.odci.gov/cia/publications/factbook/print/us.html. Accessed September 2004.

Daily, Gretchen C. and Paul R. Ehrlich. 1992. Population, Sustainability, and Earth's Carrying Capacity: A Framework for Estimating Population Sizes and Lifestyles that Could Be Sustained without Undermining Future Generations. Available online from http://dieoff.org/page112.htm. Accessed September 2004.

Daly, Herman E. and John B.Cobb, Jr. 1989. *For the Common Good.* Boston, MA: Beacon Press.

Gujarati, Damodar N. 1995. *Basic Econometrics*. 3rd. ed. New York: McGraw-Hill, Inc.

Hamilton, C. and J. Perry. 1962. A short method for projecting population by age from one decennial census to another. *Social Forces*, 41: 163 – 170.

Kentucky State Data Center. 2003. Kentucky Population Estimates: Intercensal Population Estimates, 1991 – 1999. Available online from http://ksdc.louisville.edu/kpr/popest/ice9000.xls. Accessed September 2004.

Kentucky State Data Center. 2003. Kentucky Population Projections. Available online from http://ksdc.louisville.edu/Projections2003.htm. Accessed September 2004.

Kentucky State Data Center. 2003. Kentucky Population Projections: Assumptions and Methodology. Available online from http://ksdc.louisville.edu/kpr/pro/assumptions.htm. Accessed September 2004.

Kentucky State Data Center. 2003. Kentucky Population Projections: Middle Projection Series. Available online from http://ksdc.louisville.edu/kpr/pro/Middle_Series.xls. Accessed September 2004.

Kentucky State Data Center. 2003. Kentucky Population Projections: Populations 1990 – 2000, with Middle, Low, and High Projections 2005 – 2030. Available online from http://ksdc.louisville.edu/kpr/pro/Summary_Table.xls. Accessed September 2004.

Kentucky State Data Center. 2004. Available online from http://ksdc.louisville.edu/. Accessed September 2004.

Keyfitz N. 1968. *An Introduction to the Mathematics of Population.* Reading, MA: Addison-Wesley.

Klosterman, Richard E. 1990. *Community Analysis and Planning Techniques*. Savage, MD: Rowman and Littlefield Publishers, Inc.

National Center for Health Statistics. 1999. Births and deaths: preliminary data for 1998. *National Vital Statistics Report*, 47(25). Available online from http://www.cdc.gov/nchs/data/nvsr/nvsr47/nvs47_25.pdf. Accessed September 2004.

National Center for Health Statistics. 2001. Births: final data for 2000. *National Vital Statistics Report*, 50(5). Available online from http://www.cdc.gov/nchs/data/nvsr/ nvsr50/nvsr50_05.pdf. Accessed September 2004.

National Center for Health Statistics. 2002. United States life tables, 2000. *National Vital*

Statistics Report, 51(3), Available online from http://www.cdc.gov/nchs/data/nvsr/nvsr51/nvsr51_03.pdf. Accessed September 2004.

National Center for Health Statistics. 2002. Births: preliminary data for 2002. *National Vital Statistics Report*, 51(11). Available online from http://www.cdc.gov/nchs/products/pubs/pubd/nvsr/51/51-12.htm. Accessed September 2004.

National Center for Health Statistics. 2004. Available online from http://www.cdc.gov/ nchs/. Accessed September 2004.

Smith, Stanley K., Jeff Tayman and David A. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Smith, Stanley K. and Stefan Rayer. 2004. *Florida Population Studies*, Bureau of Economic and Business Research (BEBR), 37(2). Bulletin 138.

U.S. Census Bureau. 2000. Fertility of American Women: June 2000 (P20 – 543RV). Available online from http://www.census.gov/prod/2001pubs/p20 – 543rv.pdf. Accessed September 2004.

U.S. Census Bureau. 2000. Migration by Sex and Age for the Population 5 Years and Over for the United States, Regions, States, and Puerto Rico: 2000 (PHC-T-23). Available online from http://www.census.gov/population/www/cen2000/phc-t23.html. Accessed September 2004.

U.S. Census Bureau. 2000. Summary File 1(SF1). Available online from http://factfinder.census.gov/servlet/DTGeoSearchByListServlet?ds_name=DEC_2000_SF1_U& state=dt. Accessed September 2004.

U.S. Census Bureau. 2002. State Population Estimates by Selected Age Categories and Sex: July 1, 2002. Available online from http://www.census.gov/popest/archives/2000s/vintage_2002/ST-EST2002-ASRO-01.html. Accessed September 2004.

U.S. Census Bureau. 2003. Domestic Migration across Regions, Divisions, and States: 1995 – 2000. Available online from http://www.census.gov/prod/2003pubs/censr-7.pdf. Accessed September 2004.

U.S. Census Bureau. 2004. Metropolitan and Micropolitan Statistical Area Definitions. Available online from http://www.census.gov/popest/estimates.php. Accessed September 2004.

U.S. Census Bureau. 2004. Population Estimates. Available online from http://www. census.gov/popest/estimates.php. Accessed September 2004.

U.S. Census Bureau. 2004. Population Estimates Terms and Definitions. Available online from http://www.census.gov/popest/topics/terms/. Accessed September 2004.

U.S. Census Bureau. 2004. Schedule of Population and Household Projection Releases. Available online from http://www.census.gov/population/www/projections/projsched.html. Accessed September 2004.

U.S. Census Bureau. 2004. State Population Projections. Available online from http: //www.census.gov/population/www/projections/stproj.html. Accessed September 2004.

Whelpton. 1928. Population of the United States, 1925 to 1975. American Journal of Sociology 34: 253 – 270.