Technical Appendix

A.1 Population Estimates

A population estimate is the determination of the size or the characteristics of a population at a current or past date in the absence of census data for the same date. In the United Sates, an estimate usually is made on a de jure basis, which means that people are estimated where they usually reside. This makes sense because the U.S. census is conducted on a de jure basis. However, there also is a need for estimating the de facto population of a given place at a given time and researchers have developed such estimates (Swanson and Pol 2005, 2008; Swanson and Tayman 2011). These estimates include vacationers (of interest, for example, to the casino industry in Las Vegas and the Hawaii Visitors Bureau), migratory workers (of interest, for example, to health care, school, and other social service providers), and the people who work in the central business district of a large city each day, but leave it largely vacant in the evenings (of interest to the San Francisco City Planning Office, for example). While estimates of de facto populations are of great interest, they are very difficult to make in the United States because of the lack of census benchmarks (Cook 1996; Smith 1994; Swanson and Tayman 2011). An estimate can be prepared for a nation or a sub-national area such as a state, county, city, town, or census tract. An estimate also can be prepared for groups of sub-national areas, groups of nations, or even the world as a whole. The principal demographic characteristics for which an estimate is made include age and sex. However, in multiracial and multi-ethnic countries such as the United States and Canada, an estimate might be done not only by age and sex, but also by race and ethnicity. An estimate also can be made of social and economic sub-groups of the population, households, and families.

The term "population estimate" is frequently used in the public domain to refer to the determination of the size or the characteristics of a population at a future date. However, most demographers prefer to use the term projection when talking about the possible size and characteristics of a population in the future. In developing a portrait of a given population in the future, it is not uncommon for a series of projections to be made that incorporate a range of plausible assumptions (e.g., expected trends in fertility, mortality, and migration). However, when one of these

D.A. Swanson, *The Washington State Census Board and Its Demographic Legacy*, SpringerBriefs in Population Studies,
DOI 10.1007/978-3-319-25948-2

projections is selected as representing the most likely future, it then becomes a forecast for the population in question. As opposed to a projection or a forecast, a population estimate is concerned with either the present or the past, but not the future (Smith Tayman and Swanson 2013: 3–4). Thus, it is useful to make the following distinctions among the terms "estimate," "projection," and "forecast."

Estimate—A calculation of a current or past population, typically based on symptomatic indicators of population change.

Projection—The numerical outcome of a particular set of assumptions regarding future population trends.

Forecast—The projection deemed most accurate for the purpose of predicting future population.

Virtually all methods of population estimation can be categorized into one or the other of two traditions: (1) demographic (Bryan 2004a); and (2) statistical (Kordos 2000; Platek et al. 1987; and Rao 2003). Demographic methods are used to develop estimates of a total population as well as its ascribed characteristics, age, race, and sex. Statistical methods are largely used to estimate the achieved characteristics of a population, and include, for example, educational attainment, employment status, income, and marital status. As is the case in the national statistical agencies of other countries, the US Census Bureau produces estimates using both of these traditions, demographic and statistical.

Demographers and statisticians have developed a wide range of estimation methods designed to meet different information needs at varying levels of accuracy and cost. As noted earlier, for the most part they are based on the concept of a de jure population although there are exceptions (Swanson and Pol 2005). The methods can be roughly placed into three categories: (1) analytical and statistical models that use data symptomatic of population and its changes; (2) mathematical models that use historical census data; and (3) sample surveys. Methods falling into the first category have generally been developed by and for applied demographers, many of whom work for national, state, and local governments. Methods falling into the second category have generally developed by and for academic demographers, many of whom work at universities and research institutes. The methods falling into the third category have generally been developed by and for statisticians and survey research scientists, but they also are widely used by demographers. Not surprisingly, there also are techniques that combine methods from two or even all three categories.

Population estimation methods also can be identified along a temporal dimension: (1) inter-censal estimates, which refer to a date between two census counts and usually take the results of both counts into consideration; (2) post-censal estimates, which refer to a date subsequent to the latest census count and usually take into account one or more previous census counts; and (3) pre-censal estimates, which refer to a date prior to a census count, but usually take into account one or more subsequent census counts. This temporal classification is useful because different methods are typically employed in the development of inter-censal, post-censal, and pre-censal estimates (Bryan 2004b).

Among survey statisticians, the demographer's definition of an estimate is generally termed an "indirect estimate" because unlike a sample survey, the data used to construct a demographic estimate do not directly represent the phenomenon of interest (Swanson and Stephan 2004: 758 and 763). In the context of the present work, the definition of an estimate found in the demographic tradition is used.

There are other ways to classify estimation methods. John Long (1993), for example, categorizes them generally into two types: (1) "flow" methods; and (2) "stock" methods. Flow methods are also known as component methods, because they require estimation of each component of population change (births, deaths, and migrants) since the last census. Stock methods relate changes in population size since the last census to changes in other measured variables: the number of housing units, automobile registrations, total number of deaths (and births), and tax returns. Long (1993) also notes that stock and flow methods may be used in combination. Popoff and Judson (2004: 603), make the following useful distinctions between stocks and flows: "…stock data are the numbers of persons at a given date, classified by various characteristics…(and) are recorded from censuses….flow data are the collection of or summation of events. At the most basic level this includes births, deaths, and migration…." This distinction is useful for purposes of this paper because, as is discussed later in this section, there are population estimations methods that solely rely on "stock" data while others rely on a combination of "stocks" and "flows."

Finally, it is useful here to consider micro data and aggregated data in the context of population estimation methods. "Micro data" means records for individual persons. These records are often linked by relationships to form family and household records and the term "micro data" refers to these linked records as well. The "Public Use Microdata Sample" (PUMS) is such a file (Swanson and Stephan, 2004: 772). "Aggregated data" are summations of records of individuals (families and households) such as one would find in a table. The aggregations are often done to specific geographic areas, but they can also be done for types of people across different geographies. The life table constructed by Kintner and Swanson (1994) for retirees of General Motors is an example of such an aggregation.

The development of methods of population estimation roughly corresponds to the development of censuses and vital statistics registries. For example, in the late 17th century, John Graunt estimated the population of London and then of the whole of England and Wales using what today is known as a censal-ratio method (Devlin 2008: 93–94). Not long afterward, the French mathematician, Laplace, also used a censal-ratio method in combination with recorded births and a population sample to estimate the population of France (Stigler 1986: 163–164). However, methodological development really only took off in the late 1930s and early 1940s, fueled in large part by the need for low-cost and timely information generated by the great depression of the 1930s and World War II (Bryan 2004a; Eldridge 1947; Hauser and Tepping 1944; Shryock 1938; Shryock and Lawrence 1949). In the United States, the Census Bureau played a major role in this effort, but it was not alone. As described in this book the Washington State Census Board developed a comprehensive program of annual population determinations based on estimation methods that are still used today (Swanson and Pol 2005, 2008). Around this same time, demographers also began developing estimation methods for what were then called “underdeveloped countries,” (Brass 1968; Chandrasekaran and Deming 1949; Popoff and Judson 2004; United Nations 1969) and the use of sample surveys as a substitute for complete census counts took hold (Bryan 2004a).

Today, population estimates are ubiquitous. They are done around the world by a host of governmental and non-governmental entities, as well as individual consultants (Bryan 2004a; Siegel 2002; Swanson and Pol 2008). The widespread availability of data, methods, and technology has made is possible for many people not only to develop estimates, but to do so more quickly and less expensively than has ever been done before. This trend is not likely to abate, but it carries certain costs in that estimates may both be made and used with little or no understanding of the issues involved, what constitutes good estimates, and how to identify them.

A.2 The Housing Unit Method (HUM)

The HUM is designed to generate estimates of the total population by focusing on the population residing in households. As such, it inherently fits within the demographic tradition. However, while the HUM is inherently demographic in nature, two key HUM elements, Persons Per Household (PPH) and Vacancy Rate (VR) are generated using methods that fit within the statistical tradition. Thus, both traditions are covered in discussing the HUM. Given that the HUM is aimed at the population residing in households, it is easy to see that is used to generate estimates of the total “De jure” population. This, of course, is the definition of population used by the US Census Bureau, which is based on place of “usual residence” (Cook 1996; Wilmoth 2004).

One of the first times that the HUM is mentioned in the academic literature is found in an article by Starsinic and Zitter (1968), who found that it made a “… surprisingly strong showing…” and that “…it may be worthwhile to devote considerably more effort to refining the input data for estimating the number of households in addition to dealing with the problem of deriving current estimates on the size of households” (Starsinic and Zitter 1968: 484). The article mentions work by Carl Frisén (1958) on the HUM in the 1950s for the California Department of Finance. The work by Frisén involved testing methods of population estimation against special censuses done by the state of California during the 1950s. Earlier work along these lines by was reported by Frisén (1951) when he was at San Jose State University.

However, the HUM was used even before 1950. It was used in the 1940s under the auspices of the Washington State Census Board (Lowe 2009). As detailed elsewhere in this book, the Board was abolished in 1967, and its operations transferred to the Washington State government. It exists in the state’s Office of Financial Management (Lowe 2009). Washington’s use of the HUM is done in conjunction with census counts that allows cities and towns to conduct a special 'headcount' census when disagreements over estimates arise (Washington Office of Financial Management 1978, 2007). These census counts are conducted in accordance with residency and housing definitions used by the Census Bureau with training assistance and supervision (including auditing) from the Washington Office of Financial Management. In 1981, the Washington system of municipal population estimation was adapted by the state of Alaska (Alaska Department of Community and Regional Affairs 1981a, b; Alaska Department of Labor 1981, 1982; Swanson Baker and Van Patten 1983). Today, the HUM is arguably the most commonly used method of population estimation in the United States (Bryan 2004a).

The Housing Unit Method (HUM) is a "stock" method that describes a basic identity in the same way that the balancing equation does (Bryan 2004a). In the case of the HUM, this identity is usually given as

$$P = H * (1 - VR) * PPH + GQ \tag{A.1}$$

where

P = Population
H = Housing units,
VR= Vacancy Rate (Proportion Vacant),
PPH = Average number of persons per household, and
GQ = Population residing in "group quarters" and the homeless.

The HUM equation can be expressed in less detail (i.e., P = HH * PPH + GQ, where HH = H * (1 − VR), Smith and Cody 2004: 2) or more detail - by structure type, for example (Swanson et al. 1983). It also can be used in combination with sample data, which opens the door to developing measures of statistical uncertainty for the estimates so produced (Roe et al. 1992).

The HUM is based on the assumption that virtually everyone lives in some type of housing structure. It is generally accepted that the HUM is the most commonly used method for making small area population estimates in the United States (Byerly 1990; Smith 1986; Smith et al. 2002). Because of how data are collected, the HUM has not been a method that could be used for all sub-national areas and the nation as a whole until recently. However, with the continuous "Master Address File," it has now emerged as a method that can be used by the US Census Bureau for all sub-national areas and the nation as a whole (Wang 1999).

Key issues in making the HUM work are: (1) the stock of housing units in the area to be estimated; (2) the vacancy rate for this stock (which when subtracted from 1.00 becomes the occupancy rate); (3) the average number of persons per household for this stock; and (4) the number of people in the area not residing in the housing stock (i.e., those in group quarters and the homeless).

A.3 Censal Ratio Method

The censal-ratio method can be implemented in several different ways. The most basic approach is to use relationships between symptomatic indicators and population counts in census years to estimate populations in non-census years and applying these relationships to symptomatic indictors available in the years for which estimates are desired. The general form of this approach is as follows.

$$R_{i,j,t} = S_{i,j,t} / P_{i,t} \tag{A.2a}$$

where

R = Censal-ratio
P = population
S = symptomatic indicator
j = indicator $(1 \le j \le k)$
i = subarea $(1 \le j \le n)$
t = year of the most recent census

Once a censal-ratio is constructed, a population estimate for time t+k is developed by dividing the t+k value of the symptomatic indicator ($S_{i,j,t+k}$) by the ratio ($R_{i,j,t}$)to yield an estimate of $P_{i,t+k}$:

$$\hat{P}_{i,t+k} = S_{i,j,t+k} / R_{i,j,t} \tag{A.2b}$$

If area i has a parent area for which an independently-derived population estimate is available, it is common is to effect a final "control" so that the sum of the i subarea population estimates is equal to the independently estimated population for the parent of these i subareas, $\sum \mathbf{P_{i,t+k}}$, which is accomplished as follows:

$$\hat{P}_{i,t+k} = (\hat{P}_{i,t+k} / \sum \hat{P}_{i,t+k}) * \left(\sum \mathbf{P_{i,t+k}}\right) \tag{A.2c}$$

It should be noted that as long as the algebra yields an estimate of P_i at time t + k, it is immaterial if $R_{i,j,t} = P_{i,t} / S_{i,j,t}$ or if $R_{i,j,t} = S_{i,j,t} / P_{i,t}$. In the case of the latter version, Equation (A.2a) and (A.2b) become, respectively

$$R_{i,j,t} = P_{i,t} / S_{i,j,t} \tag{A.2d}$$

$$\hat{P}_{i,t+k} = (R_{i,j,t}) / (S_{i,j,t+k}) \tag{A.2e}$$

One advantage of using Eqs. (A.2a) and (A.2b) over (A.2d) and (A.2c) is that resulting ratio of interest is easier to interpret. If one uses deaths as the symptomatic indicator, then the ratio is the crude death rate. Similarly, if one uses births, the resulting ratio is the crude birth rate.

A.4 Ratio-Correlation Method

The most common regression-based approach data to estimating the total population of a given area is the ratio-correlation method. It is an extension of the censal l ratio method, one that can incorporate multiple ratios and accommodate changes in the ratios over time. It was introduced and tested by Schmitt and Crosetti (1954) and again tested by Crosetti and Schmitt (1956). This multiple regression method involves relating between changes in several variables known as symptomatic indicators on the one hand to population changes on the other hand. The symptomatic indicators that are used reflect population change. Examples of symptomatic variables that have been used for this purpose are births, deaths, school enrollment, tax returns, motor vehicle registrations, employment data, and registered voters. The ratio-correlation method is used where a set of areas (e.g., counties) are structured into a geographical hierarchy (e.g. the populations of counties within a given state sum to the total state population). It proceeds in two steps. The first is the construction of the model and the second is its implementation —actually using it to create estimates for given years.

Because the method looks at change, population data from two successive censuses are needed to construct the model along with data for the same years representing the symptomatic indicators. During its implementation step the ratio-correlation method requires symptomatic data representing the year for which an estimate is desired and an estimate of the population for the highest level of geography (e.g., the state as a whole) that is independent of the ratio-correlation model.

The ratio-correlation method expresses the relationship between (1) the change over the previous inter-censal period (e.g., 1990 to 2000) in an area's share (e.g., a given county) of the total for the parent area (e.g., the state as a whole) for several symptomatic series and (2) the change in an area's share of the population of the parent area. The method can be employed to make estimates for either the primary or secondary political, administrative and statistical divisions of a country (Bryan 2004).

In general terms, the ratio-correlation model is formally described as follows (Swanson and Beck 1994):

$$P_{i,t} = a_0 + \sum (b_j) * S_{i,j,t} + \varepsilon_i \tag{A.3a}$$

where

a_0 = the intercept term to be estimated
b_j = the regression coefficient to be estimated
ε_i = the error term
j = symptomatic indicator ($1 \le j \le k$)
i = subarea ($1 \le j \le n$)
t = year of the most recent census

and

$$P_{i,t} = \left(P_{i,t} \Big/ \sum P_{i,t}\right) \Big/ \left(P_{i,t-z} \Big/ \sum P_{i,t-z}\right) \quad \text{(A.3b)}$$

$$S_{i,t} = \left(S_{i,t} \Big/ \sum S_{i,t}\right) \Big/ \left(S_{i,t-z} \Big/ \sum S_{i,t-z}\right) \quad \text{(A.3c)}$$

where

z = number of years between each census for which data are used to construct the model
p = population
s = symptomatic indicator

Once a ratio-correlation model is constructed, a set of population estimates for time t + k is developed in a series of six steps. First, $\left(S_{i,t+k}/\sum S_{i,t+k}\right)_j$ is substituted into the numerator of the right side of Eq. (A.3c) for each symptomatic indicator j and $\left(S_{i,t}/\sum S_{i,t}\right)_j$ into the denominator of the right side of Eq. (A.3c) for each symptomatic indicator j, which yields $S_{i,j,t+k}$ Second, the updated model with the preceding substitution of symptomatic data for time t + k is used to estimate $P_{i,t+k}$. Third, $(\left(P_{i,t}/\sum P_{i,t}\right))$ is substituted into the denominator of $P_{i,t+k}$, which yields $P_{i,t+k} = \left(P_{i,t+k}/\sum \mathbf{P_{i,t+k}}\right)/\left(P_{i,t}/\sum P_{i,t}\right)$, where $\sum \mathbf{P_{i,t+k}})$ represents the independently estimated population of the "parent" area of the i subareas for time t + k (Note that this estimate is given in boldface and is done by a method exogenous to the ratio-correlation model (e.g., a component method)). Fifth, since $P_{i,t+k}$, $\left(P_{i,t}/\sum P_{i,t}\right)$ and $\sum \mathbf{P_{it+k}}$ are all known values, the equation $P_{i,t+k} = \left(P_{i,t+k}/\sum \mathbf{P_{i,t+k}}\right)/\left(P_{i,t}/\sum P_{i,t}\right)$ is manipulated to yield an estimate of the population of area i at time t + k:

$$\left(P_{i,t+k}\right) * \left(P_{i,t} \Big/ \sum P_{i,t}\right) * \left(\sum \mathbf{P_{i,t+k}}\right) = P_{i,t+k} \quad \text{(A.3d)}$$

As Eq. (A.3d) shows, it is important to remember that an independent estimate of the population for the "parent" geography $\left(\sum \mathbf{P_{i,t+k}}\right)$ of the i subarea is required when using the ratio-correlation model to generate population estimates. The sixth and final step is to effect a final "control" so that the sum of the i subarea population estimates is equal to the independently estimated population for the parent of these i subareas: $\sum P_{i,t+k} = \sum \mathbf{P_{i,t+k}}$, which is accomplished as follows:

$$P_{i,t+k} = \left(P_{i,t+k} \Big/ \sum P_{i,t+k}\right) * \left(\sum \mathbf{P_{i,t+k}}\right). \quad \text{(A.3e)}$$

It should be clear from the preceding definitions that we are focusing on the ratio-correlation method as a means of developing post-censal estimates. However, it can be used to develop inter-censal estimates. It also could be run in reverse to estimate "historical" populations.

There are variations on the standard form discussed here. They include the rate-correlation model (Swanson and Tedrow 1984) and the difference-correlation model (Schmitt and Grier 1966).

A.5 Component Methods

There are several methods of population estimation that belong to the "component" family. All of the component methods are based on the fundamental demographic equation:

$$P_{i,t+k} = P_{i,t} + B_i - D_i + I_i - O_i \tag{A.4a}$$

where

$P_{i,t}$ = Population of area i at time t (the launch date)
$P_{i,t+k}$ = Population of area i at time t + k (the estimate date)
B_i = Births in area i between time t and t + k
D_i = Deaths in area i between time t and t + k
I_i = In-migrants in area i between time t and t + k
O_i = Out-migrants in area i between time t and t + k

This deceptively simple equation can be displayed in a number of forms (Hoque 2010; Murdock et al. 1995; Zitter and Shryock 1964). For example, it is common to combine in-migrants and out-migrants into net number of migrants and use the fundamental equation to estimate net migration between two censuses one taken at time t and the other at time t + k:

$$N_i = P_{i,t+k} - P_{i,t} - B_i + D_i \tag{A.4b}$$

where

$P_{i,t}$ = Population of area i at time t
$P_{i,t+k}$ = Population of area i at time t + k
B_i = Births in area i between time t and t + k
D_i = Deaths in area i between time t and t + k

Ni = $I_i - O_i$ =
(In-migrants to area i between time t and t + k) − (Out-migrants from area i between time t and t + k)

To be exactly true, the fundamental equation must apply to a defined population (e.g., the resident population) of a fixed area i and there must be no measurement errors. For example, if we are using it to estimate the resident population of area i at time t + k, then all births and deaths used must be to the resident population of area i between time t and time t + k while all in- and out-migrants during the same period also apply to this same resident population and $P_{i,t}$ is measured without error.

The fundamental equation can be applied to age, sex, race, and ethnic segments of the population. In the case of an age group the age specification of the group changes over the period. For example, if t is 10 years, then one should compare age x at time 0 with x + 10 at time t. Put another way, this age group is a "cohort" that is followed over time. In conjunction with age groups, and the use of future fertility, mortality and net migration rates, the expanded version of the fundamental equation can be used to make both estimates and projections. This is known as the "cohort-component method" (Smith et al. 2013), where "cohort" is defined as before and "component" is used to refer to the three components of population change, fertility, mortality, and migration.

All of the component methods generally employ counts of births and deaths because they are generally available every year from vital statistics records while migration data are only available in countries with well-maintained population registers (e.g., Finland). They tend to vary in how the migration component is estimated. Two examples follow, "Component Method II" and the "Cohort-Component Method."

A.5.1 Component Method II

Component Method II (CM II) is based on an estimate of net migration that finds the difference between a current estimate of school-age children (e.g., time = t + k) in area i with the expected number "survived" from the last census (e.g., time = t) of area i and then converting the difference to a migration rate that is applied to the entire population of area i at time t. The net migration component is estimated in six steps: (1) Enrollment in selected grades (e.g. grades 2 to 8 or in grades Kindergarten to 9) at time = t + k is adjusted to approximate the population of corresponding elementary school age on the basis of the relative size of these two groups at the last census (relating local school enrollment data to a census count at time t); (2) next, the "expected" population (assuming no net migration) of elementary school age for area i for time t + k is found by "surviving" the population in the same cohort from time t (including, if necessary, births subsequent to time t) to t + k (This is usually done using survivorship probabilities found a life table that is assumed to apply to area for the period t to t + k; (3) the net migration of children of school age is estimated as the difference between the "actual" population of school age and the "expected" population of school age; (4) the estimated net migration of school-age children is converted into the estimated net migration of the remainder of the population by dividing these other population groups by the number of school age children at the time of the last census; (5) the estimated net number of migrants in each age group is then summed to obtain an estimate of the net number of migrants for the total population; and (6) in the final step, the total population is obtained by using the fundamental demographic equation: adding to the population in the last census, the net number of migrants and the number of births during the period and subtracting the number of deaths.

Where administrative records data are available on the population aged 65 years and over (e.g., in the US Medicare data), it is not uncommon to use CM II to develop an estimate of the population less than 65 years with appropriate adjustments to the six steps just described and then use the administrative records data to estimate the population age 65 years and over (Murdock et al. 1995). The two groups are added together to get an estimate of the total population in what could be termed a composite method (Bogue 1950; Bogue and Duncan 1959). There are more variations on the basic idea (Bryan 2004b; U.S. Census Bureau 2010; Zitter and Shryock 1964), but these six steps essentially describe CM II.

CM II assumes: (1) there has been no change since the last census in the ratio of the population of elementary school age to the number enrolled in the elementary grades; and (2) that the ratio of the net migration rate of the total population to the migration rate of the school-age population of area i for the period t to t + k corresponds to that for the net migration of adults for this are over the same period.

It is worthwhile to note that other variations in the use of school data to estimate net migration in a component model are possible. One is the "grade-progression method," which determines the annual net migration of school-age children by comparing the number of children enrolled in, for example, grades 2 to 7 in one year with the number enrolled in grades 3 to 8 in the following year. The remaining steps in a school-progression approach are those described for CM II.

A.5.2 Cohort-Component Method

The cohort-component method was introduced by Cannan (1895), subsequently used by Bowley (1924), and later re-discovered independently by Whelpton (1928). It is the most widely used method for producing population projections. Since it is used for projections it also can be used for estimates. Whether used for projections or estimates, the basic framework is the same as shown in Equations (A.4a) and (A.4b), but with age and sex details. We only provide an overview of the cohort-component method here. Full implementation details are found in Smith et al. (2013).

The cohort-component method divides the population at time = t (the launch date) population into age-sex groups (i.e., cohorts) and accounts separately for the fertility, mortality, and migration behavior of each cohort as it passes from he launch date at time = t to the estimate data at time = t + k. The division of the population into age groups was an important methodological advance (de Gans 1999). Not only does this account for the differences in mortality, fertility, and migration rates among different age groups at a particular time, but it also allows for changes in these rates for individual cohorts as they cycle through time.

Age cohorts can be defined in a number of ways, but cohort-component models typically use either single years or 5-year groups. The oldest age group is virtually always "open-ended," usually 75+, 85+, or 90+. Age groups are typically divided by sex and are sometimes further subdivided by race, ethnicity, and other ascribed characteristics.

The cohorts are cycled through time in "intervals," where the components of change are applied to the cohorts in each interval as appropriate to bring them forward in time from the launch date. It is customary that the width of the number of years used to define the cohorts corresponds to the number of years in the temporal interval (i.e., 5-year age cohorts when the cohort-component method uses 5-year intervals).

The first step in the process is to establish the launch year (time = t) population and calculate the number of persons in it who survive to the estimation date (tie = t + k) . This is done by applying age-sex-specific survival rates to each age-sex group in the launch year population. These can be "controlled" so that the numbers they generate match reported deaths for each interval (e.g., year) up to the estimate date.

The second step is to calculate migration for each age-sex group in each interval from time = t to time = t + k. The third step is to calculate the number of births in each interval. This is usually done by applying age-specific birth rates to the female population in each age group. As was the case with the age-sex specific survival rates, these can be "controlled" so that the numbers they generate match reported deaths for each year up to the estimate date.

The fourth and final step in the process is to add the number of births (distinguishing between males and females) to the rest of the population. These calculations provide an estimate of the population by age and sex at the end of each interval. This population then serves as the starting point for the following interval. The process is repeated until the estimate date is reached.

*The discussions of methods are largely taken from Swanson and Tayman (2013).

References

Alaska Department of Community and Regional Affairs. (1981a). *The housing unit method of population estimation: A manual for municipal personnel responsible for annual population estimates.* Juneau, AK: Alaska Department of Community and Regional Affairs, Division of Local Government Assistance and Alaska Department of Labor, Research and Analysis Section.

Alaska Department of Community and Regional Affairs. (1981b). *Standards for conducting a population census in small Alaskan cities.* Juneau, AK: Alaska Department of Community and Regional Affairs, Division of Local Government Assistance.

Alaska Department of Labor. (1981). *State of Alaska census enumerator's manual*. Juneau, AK: Alaska Department of Labor, Research and Analysis Section.

Bogue, D. (1950). A technique for making extensive population estimates. *Journal of the American Statistical Association,45*, 149–163.

Bogue, D., & Duncan, B. (1959). A composite method for estimating postcensal population of small areas by age, sex, and color. *Vital Statistics Special Reports,47*(6).

Bryan, T. (2004a). Population estimates. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography, 2nd Edn* (pp. 523–560). New York, NY: Elsevier Academic Press.

Bryan, T. (2004b). Basic sources of statistics. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography, 2nd Edn* (pp. 9–41). New York, NY: Elsevier Academic Press.

Byerly, E. (1990). State and local agencies preparing population and housing estimates. *Current Population Reports, Series P-25-1063*. Washington, DC: U. S. Government Printing Office.

Cook, T. (1996). When ERPs aren't enough: A discussion of issues associated with service population estimation. *Working Paper 96/4*. Demography Section, Australian Bureau of Statistics, Belconnen, ACT, Australia.

Crosetti, A., & Schmitt,R. (1956). A method of estimating the intercensal populations of counties. *Journal of the American Statistical Association,51*(276), 587–590.

Devlin, K. (2008). *The unfinished game: Pascal, Fermat, and the seventeenth-century letter that made the world modern.* New York, NY: Perseus Publishers.

Frisén, C. (1951). *Symptomatic data in population estimates: Problems and limitations*. Paper presented at the annual meeting of the Pacific Sociological Association, Berkeley, CA (published in *Research Studies of the State College of Washington*, Vol. 19, pp. 107–110).

Frisėn, C. (1958). Effectiveness of our tools for estimating population change in small areas. In *Proceedings of the American Statistical Association, Social Statistics Section* (pp. 229–232). Alexandria, VA: American Statistical Association.

Hauser, P., & Tepping, B. (1945). Evaluation of census wartime population estimates and predicitons of postwar population prospects for metropolitan areas. *American Sociological Review, 9*(5), 473–480.

Hoque, M. N. (2010). An evaluation of small area population estimates produced by component method II, ratio-correlation, and housing unit methods. *The Open Demography Journal,3*, 18–30.

Kintner, H., & Swanson, D. (1994). Estimating vital rates from corporate databases: How long will GM's salaried retirees live?. In H. Kintner, T. Merrick, P. Morrison, & P. Voss (Eds.), *Demographics: A casebook for business and government* (pp. 265–295) . Boulder, CO: Westview Press.

Kordos, J. (Ed.). (2000). Special issue on small area estimation. *Statistics in TRANSITION: Journal of the Polish Statistical Association,* 4(4).

Long, J. (1993). An overview of population estimation methods. In D. Bogue, E. Arriaga & D. Anderton (Eds.), *Readings in population research methodology* (Vol. 5, pp. 20–1 to 20–5). Chicago: United Nations Population Funds and Social Development Center.

Lowe, T. (2009). Personal communication.

Murdock, S., Hwang, S., & Hamm, R. (1995). Component methods. In N. Rives, W. Serow, A. Lee, H. Goldsmith, & P. Voss (Eds.), *Basic methods for preparing small-area population estimates* (pp. 10–53). Madison, WI" Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin.

Platek, R., Rao, J., Sarndal, C., & Singh, M. (Eds.). (1987). *Small area statistics: An international symposium*. New York, NY: John Wiley.

Rao, J. N. K. (2003). *Small area estimation*. New York, NY: Wiley-Interscience.

Popoff, C., & Judson, D. (2004). Some methods of estimation for statistically underdeveloped areas. In J. Siegel & D. Swanson (Eds.), *The Methods and Materials of Demography, 2nd Edn* (pp. 603–641). New York, NY: Elsevier Academic Press.

Schmitt, R., & Crosetti, A. (1954). Accuracy of the ratio-correlation method for estimating poptcensal population. *Land Economics,30*(3), 279–281.

Schmitt, R., & Grier, J. (1966). A method of estimating the population of minor civil divisions. *Rural Sociology,31*, 355–361.

Shryock, H. (1938). Methods of estimating post-censal population. *American Journal of Public Health,28*, 1042–1047.

Shryock, H., & Lawrence, N. (1949). The current status of state and local population estimates in the census Bureau. *Journal of the American Statistical Association,44*(246), 157–173.

Smith, S. (1986). A review and evaluation of the housing unit method of population estimation. *Journal of the American Statistical Association,81*, 287–296.

Smith, S. K. (1994). Estimating temporary populations: The contributions of Robert C. Schmitt. *Applied Demography*, *9*(1), 4–7.

Smith, S. K., Nogle, J., & Cody, S. (2002). A regression approach to estimating the average number of persons per household. *Demography,39*(4), 697–712.

Smith, S., & Cody, S. (2004). An evaluation of population estimates in Florida: April 1, 2000. *Population Research and Policy Review,23*(1), 1–24.

Smith, S., Tayman, J., & Swanson, D. A. (2013). *A practitioner's guide to state and local population projections.* Dordrecht, The Netherlands: Springer.

Starsinic, D., & Zitter, M. (1968). Accuracy of the housing unit method in preparing population estimates for cities. *Demography,5*, 475–484.

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1990.* Cambridge, MA: Belknap Press of Harvard University Press.

Swanson, D., & Stephan, G. E. (2004). Glossary. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography, 2nd Edn* (pp. 751–778). New York, NY: Elsevier Academic Press.

Swanson, D. A., & Tedrow, L. (1984). Improving the measurement of temporal change in regression models used for county population estimates. *Demography,20*, 373–382.

Swanson, D. A., Baker, B., & Van Patten, J. (1983). *Municipal population estimation: Practical and conceptual features of the housing unit method*. Paper Presented at the 1983 Annual Meeting of the Population Association of America, Pittsburgh, PA, April 14th–16th. (Reprinted as pp. 99–105 in *Alaska Population Overview 1982*. Alaska Department of Labor, Juneau, AK.

Swanson, D., & Pol, L. (2005). Contemporary developments in applied demography within the United States. *Journal of Applied Sociology,21*(2), 26–56

Swanson, D., & Pol, L. (2008) Applied demography: Its business and public sector components. In Yi Zeng (Ed.), *The encyclopedia of life support systems, demography volume*. Oxford, England: UNESCO-EOLSS Publishers. (with L. Pol).(Online at http://www.eolss.net/).

Swanson, D. A., & Tayman, J. (2011). On estimating a de facto population and its components. *Review of Economics and Finance* (5), 17–31.

United Nations. (1969). *Methodology and evaluation of population registers and similar systems.* New York, NY: United Nations.

Wang, C. (1999). *Development of national accounting of address and housing inventory: The baseline information for post-censal population estimates*. Paper prepared for the estimates methods conference, U.S. Bureau of the Census, Federal Office Building #3, Suitland, Maryland, June 8th. (available online at http://www.census.gov/population/www/coop/popconf/paper.html)

Washington Office of Financial Management. (1978). *The housing unit method: A manual for municipal personnel responsible for making annual population estimates.* Olympia, WA: Washington Office of Financial Management.

Wilmoth, J. (2004). Population size. In J. Siegel & D. Swanson (Eds.), *The methods and materials of demography, 2nd Edn* (pp. 65–80). New York, NY: Elsevier Academic Press.

Zitter, M., & Shryock, H. S. (1964). Accuracy of methods of preparing postcensal population estimates for states and local areas. *Demography,1*, 227–241.