Demographic Techniques: Small-area Estimates and Projections

Stefan Rayer, College of Liberal Arts and Sciences, University of Florida, Gainesville, FL

Abstract

Estimates and projections for small areas are used extensively in the public and private sectors, and demand for them has been growing. Because of population size and data availability issues, estimates and projections for small areas face methodological challenges not commonly encountered at larger geographical scales. This chapter defines "smallness" for estimation and projection purposes; describes the types and sources of data used; discusses the methods for estimating and projecting small-area populations; and assesses recent methodological developments, in particular the impact of GIS and spatial techniques.

Final formatted version published in *International Encyclopedia of the Social & Behavioral Sciences, 2nd Edition*, Volume 6, pp 162–169. http://dx.doi.org/10.1016/B978-0-08-097086-8.31015-7

Demographic Techniques: Small-area Estimates and Projections

Population estimates and projections play a critical role in market analysis, facility planning, environmental planning, and the allocation of public and private funds. Estimates and projections for small areas, in particular, are used extensively in the public and private sectors, and demand for them has been growing. Because of population size and data availability issues, estimates and projections for small areas face methodological challenges not commonly encountered at larger geographical scales. While most population estimation and projection techniques can be used at any scale, the focus of this chapter is on those approaches that are particularly suited for small areas.

The chapter is organized as follows: After a brief description of what distinguishes population estimates from projections, "smallness" is defined and put in context. This is followed by a discussion of the data and sources typically used when preparing small-area estimates and projections. Next, various types of estimates are defined. The central sections address in more detail the methods used for small-area estimation and projection. The chapter concludes assessing recent methodological developments, in particular the impact of geographic information system (GIS) and spatial techniques.

Estimates, Projections, and Forecasts

Demographers consider information about a past or present population not based on a census or population register an *estimate*, while information about the future is referred to as a *projection* or a *forecast*. The terms projection and forecast are often used interchangeably, but can be differentiated according to the expected likelihood of their outcomes: a projection is a numerical outcome of a particular set of assumptions regarding the future population, a forecast is the projection believed most likely to provide an accurate prediction of the future population (George, Smith, Swanson and Tayman, 2004: 561). In addition to the time dimension, a key distinction between estimates and projections involves the type of data used. Since projections refer to the size of the population at a future point in time, they cannot be based on actual data comprising the components of population change; rather, they must be based on the extension of either current or expected population trends into the future (Raymondo, 1992).

Smallness

Small area demography refers to demographic applications and analyses executed at local and regional scales (Smith and Morrison, 2005). While no universally accepted definition exists for what constitutes a small area for population estimation and projection purposes, it commonly involves administrative units below the national and state level. Examples include counties, cities, municipalities, townships, wards, local government areas, postal areas, school districts, census tracts, and census blocks; these can vary greatly in area and population size, from less than one to thousands of square kilometers, and from a handful of residents (or none) to over a million. Alternatively, a small area can refer to any subpopulation or domain for which direct estimates of adequate precision cannot be produced (Rao, 2003). Subpopulations or domains are subsets of larger populations defined by criteria other than geographic affiliation such as age, sex, income, educational level, health status, etc. Smallness can thus be understood in terms of size or in terms of data availability (Alho, 2001: 3484).

Small-area analyses face a number of unique challenges not commonly encountered at larger geographic scales: (1) the boundaries of small areas often change over time making time series analyses challenging; (2) many types of data, especially those covering more detailed population characteristics, are not tabulated for smaller areas, necessitating the use of proxy variables; (3) because of a paucity of data, there are often no discernible past patterns of change that can serve as a basis for estimation or projection, which may require the application of model rates based on areas for which data are available but which may not be directly comparable; (4) even when data are available for small areas, they may be less reliable because of smaller sample sizes and greater sampling variability; and (5) location-specific factors such as institutional populations, seasonal populations, facility closings, or changes in zoning have a greater impact on population changes in small areas (Smith and Morrison, 2005; Murdock, Cline, and Zey, 2012). Because of these challenges, population estimates and projections for small areas necessitate different sets of tools than those applied for larger areas such as states and nations.

Data and Sources

Two general categories of data can be distinguished: *direct data* and *indirect* or *symptomatic data* (Bryan, 2004b: 526). The former measure population and population change directly and can be obtained from censuses, administrative records, and surveys. Indirect data provide information indirectly related to, or symptomatic of, the population being estimated or projected; examples include school enrollment, tax returns, vital statistics, employment statistics, voter registration, electrical hook-ups, and housing counts. There is no clear dividing line between the two categories of data; the same data can be direct for one type of estimate and indirect for another. For example, birth and death counts are direct data when used to estimate natural change in a population but indirect data when used in the censal-ratio method (see below) to estimate total population (Swanson and Tayman, 2012: 43). The usefulness of indirect data depends on the extent to which factors other than population size and distribution influence them.

The primary data sources used in small area demography include censuses, administrative records, and surveys (Smith and Morrison, 2005). *Censuses* generally provide accurate and comprehensive data, regular repetition, fine geographic detail, and relative ease of data access; potential downsides include cost, enumeration difficulties, infrequent updates, falling response rates, differential underenumeration, and concerns about disclosure control (Martin, 2006). While censuses are still generally considered the "gold standard," in many countries there has been a movement towards greater utilization and application of administrative data sources and sample surveys to provide more timely data.

Administrative records, which are collected by federal, state, and local government agencies for registration, licensing, and program administration purposes, can provide ongoing information on a variety of demographic events and characteristics (Smith and Morrison, 2005). The most important administrative records for small area estimates and projections are vital statistics (especially birth and death data), which tend to be widely available and generally quite accurate, at least in the developed countries, and which form crucial inputs in many estimation and projection models. Other sources of administrative records come from a variety of fields and cover data related to the economy, education, health, social services, safety and security, community resources and participation, housing, and the environment (Coulton, 2008).

A special case of administrative records are *population registers* in which population characteristics are continually recorded. Population registers can be divided into *universal registers*, which attempt to include the entire population, and *partial registers*, which are established for specific administrative purposes and cover only those persons directly affected by the particular program (Bryan, 2004a: 31–35). Universal registers are maintained by the Scandinavian countries, the Netherlands, Japan, and a few other countries; partial registers are more common and include social insurance and welfare, military service, voter, school enrollment, and judicial data. A continuously updated enhanced master area file (EMAF) has been proposed for the United States; it has the potential to deliver timely, cost-effective, and precise population estimates even for small geographical units, but a number of challenges – ranging from confidentiality and privacy, up-front costs, to accuracy and technical issues – must be overcome for the system to become established (Swanson and McKibben, 2010).To be most useful, administrative records and population registers must be accurate and up-to-date, allow for linking from one source to another via a personal identification number, and be embedded in a supportive legislative framework (Smith and Morrison, 2005).

The third major source of data for small area estimates and projections is *surveys*. While surveys vary, their central features are the use of a fixed design, the collection of data in standardized form from individuals or organizations such as schools or businesses, and the selection of representative samples from known populations (Robson, 2011: 238). Surveys are often employed to collect data on variables not covered in a census or by administrative records. Surveys come from governmental/official statistics, academic/social research, and commercial/advertising/market research (O'Muircheartaigh, 1997: 1–2). Small sample sizes often limit the usefulness of survey estimates for small areas. This can be the case even for surveys designed specifically to provide accurate and timely demographic, social, and economic data on an ongoing basis for large and small areas, such as the American Community Survey in the United States (Swanson and Hough Jr., 2012). To combat the challenges posed by small sizes, small area estimation techniques have been developed that apply indirect estimators which "borrow strength" by using values of the variable of interest from related areas and/or time periods, thus increasing "effective" sample sizes (Rao, 2003: 2).

Types of Population Estimates

Population estimates can be divided – based on their time reference and method of derivation – into *inter-censal estimates*, which relate to a date between two censuses and take the results of these censuses into account, and *post-censal estimates*, which relate to a date following a census that take that census into account, but not later censuses (Bryan, 2004b: 523). The former can be regarded as interpolations, the latter as extrapolations. Although post-censal estimates are sometimes made with extrapolative techniques, more commonly symptomatic indicators of population change are applied (see below). Estimates can also be made for dates prior to census taking (*pre-censal*), which are of interest to historical demographers in particular. The estimation methods discussed in the following section are primarily used for making post-censal estimates, which are the most common type of population estimation. Inter-censal and pre-censal estimates can be made with some of these methods but usually require a different approach

(Swanson and Tayman, 2012: 331-355). Estimates can further be divided whether they are made for a legally resident *de jure* population or for a physically present *de facto* population. Most population estimates follow a de jure definition, and the methods described below are primarily applicable to them. De facto population estimates, which are useful for estimating daytime, visitor, seasonal, homeless, and disaster-impacted populations, for the most part utilize different sets of techniques (Swanson and Tayman, 2012: 313–330).

Methods for Small Area Population Estimates

The type and quality of data available are crucial determining factors when choosing a method for population estimation (Bryan, 2004b: 526). Estimation accuracy, and how different methods account for uncertainty, are important factors to consider; additional criteria include the provision of necessary detail, face validity, plausibility, costs of production, timeliness, and ease of application and explanation (Swanson and Tayman, 2012: 267–302). There are various ways to classify estimation methods for small areas. A recently developed classification scheme by Swanson and Tayman (2012: 106–107) combined earlier approaches and proposed the following categories: extrapolation, ratio, symptomatic, regression, component, sample based, and other methods. It should be noted that although the estimation methods discussed below have varying data requirements and can be applied across different areas, many less developed countries have underdeveloped statistical systems and are better served by indirect estimation methods which utilize model life tables and model stable populations (Popoff and Judson, 2004). For an overview of data and methods used for at-risk subnational population estimation see National Research Council (2007).

Extrapolation techniques, which range from simple linear change to complex ARIMA models, rely solely on the pattern of past population changes to estimate post-censal population, and assume that trends in the post-censal period will be similar to historical trends. Simple extrapolation models are most useful for post-censal estimates close to the last census, when resources are limited, and for very small areas or demographic subgroups; complex models allow the construction of probabilistic intervals around the estimates but are more difficult to implement and not necessarily more accurate (Swanson and Tayman, 2012: 115–127). *Ratio extrapolation methods* such as share-of-growth, shift-share, and constant-share, express the population of a subgroup as a proportion of a larger population. Like simple extrapolation techniques, they have small data requirements and are easy to apply, but share the general shortcomings of extrapolation methods in that they do not account for differences in demographic characteristics or the components of growth, and they ignore potentially relevant information related to post-censal population changes (Swanson and Tayman, 2012: 127–135).

Symptomatic estimation techniques include the housing unit and censal-ratio methods. The *housing unit method* is one of the most widely used techniques for making small area population estimates (Bryan, 2004b: 550). It relies on the assumption that nearly everyone in a population lives in some type of housing structure. The housing unit method calculates the population of an area as equal to the number of occupied housing units (households) times the average number of persons per household plus the number of persons living in group quarters (e.g. prisons, college dormitories, military barracks, nursing homes). The number of households can be estimated using measures of construction activity such as building permits or certificates of occupancy, using utility data such as a residential electric or telephone customers, from property tax records, and from aerial photographs; the average number of persons per household can be taken from the most recent census, extrapolated as a trend from the two most recent censuses, or estimated using post-censal data in combination with data from the last census; and the group quarters population can be obtained either directly from the facilities or from a past census (Smith, 1986). The housing unit method has a long and successful track record, is flexible in terms of data sources, and can be applied at most levels of geography, but it also requires a major commitment of time and resources as well as sound professional judgment to yield accurate estimates (Smith and Cody, 2004). Recent research that models housing unit change for small areas using a density-dependent growth model, as well as approaches that integrate spatial factors derived from remote sensing and GIS datasets, have shown promising results and hint at future refinements to the housing unit method (Baker, Ruan, Alcantara et al., 2008; Deng, Wu and Wang, 2010).

The *censal-ratio method*, introduced by Bogue (1950) as the "vital rates method", is another symptomatic population estimation technique used for small area estimation. It is related to the ratio extrapolation methods, but is based on ratios of symptomatic data to total population rather than proportions of national, state, or regional totals. The technique involves computing the ratio of symptomatic data to total population at the time of the last census, extrapolating the ratio to the estimate date, and dividing the estimated ratio into the value from the symptomatic series for the estimate date (Bryan, 2004b: 546–547). Symptomatic data that can be used in the censal-ratio method include birth and death statistics, school enrollment data, tax returns, number of electric, gas, or water meter accounts, number of building permits issued, bank receipts, motor vehicle registrations, and voter registration rolls. To be useful, accurate and comparable symptomatic data must be available at frequent intervals, including the census date, the annual number of cases should be high in relation to population size, and the ratio should be stable or change in a regular fashion in order to be projected to the estimate date (Bryan, 2004b: 547).

Regression techniques derive population estimates by means of symptomatic indicators of population change. The *ratio-correlation method* introduced by Schmitt and Crosetti (1954) is the most widely-used regression technique for population estimation. It involves relating changes in several symptomatic indicators to population changes – expressed in the form of ratios to totals for geographic areas – by a multiple regression equation (Bryan, 2004b: 548). The symptomatic variables that have been used are similar to those discussed above for the censal-ratio method. The *apportionment*, *ratio change*, and *additive change methods* frequently used for making small area population estimates in Great Britain can be considered as simplified versions of the ratio-correlation method (Simpson, Middleton, Diamond and Lunn, 1997). Regression techniques for population estimation have a firm foundation in statistical inference, which allows for the construction of meaningful measures of uncertainty (Swanson and Beck, 1994). The ratio-correlation method, in particular, has been used, evaluated, and refined for over half a century, giving it a large measure of quality control and contributing to its widespread use and satisfactory performance (Swanson, 2004). Regression techniques rely for their accuracy on the validity of the assumption that the relationship between the independent and dependent variables observed in the past will persist in the post-censal period; they also require judgment with regard to the reliability and consistency of coverage of the symptomatic indicators. Shortcomings include potential time lags in the availability of the symptomatic indicators; the use of multiple and differing variables, which make decomposition of error and comparability of estimates between areas difficult; and limitations in producing post-censal estimates by age and sex (Bryan, 2004b: 549–550).

Component methods are based on the demographic balancing equation in which the population at the end of the time period (here, the estimate date) is expressed as the population at the beginning of the time period (commonly the most recent census), to which the number of births and in-migrants that occurred over the time period are added, and from which the number of deaths and out-migrants that occurred over the time period are subtracted. All component methods generally employ birth and death data but vary in how migration is estimated. Of the various component methods, the *component method-II* and the *tax returns/administrative records method* developed by the U.S. Census Bureau are widely used (Murdock, Hwang, and Hamm, 1995; Starsinic, Lee, Goldsmith and Sparr, 1995). The component method-II employs changes in school enrollments to estimate migration, whereas the tax returns/administrative records method uses address changes on federal income tax returns. If migration data of sufficient quality are available, the *cohort-component method* can also be applied. It divides the population at the launch date into age-sex groups (cohorts) and accounts separately for the fertility, mortality, and migration each cohort experiences until the estimate date (Swanson and Tayman, 2012: 195–206). Although useful when estimates by age and sex are required, the cohort-component method is data- and computationally intensive and more commonly applied for population projections than for estimates. Component methods are attractive, because they specifically account for the three components of population change, births, deaths, and migration. Limitations include that they assume the continuation of historical patterns in symptomatic data in measuring migration, that they can be resource-intense and require data which may not be available for all small areas, and that they may not be appropriate for areas with substantial "special" populations (Bryan, 2004b: 556; Rives, Serow, Lee and Goldsmith, 1989: 30–31).

Sample based methods of population estimation, which include *synthetic methods*, *structure preserving estimation* (*SPREE*), the *ranked set samples method* (*RSS*), and *Bayesian methods*, are more commonly employed by statisticians than by demographers (Swanson and Tayman, 2012: 207–218); for an overview and review of small area estimation methods of this kind see Ghosh and Rao (1994), Pfeffermann (2002) and Rao (2003). Other methods that are sometimes used for creating small area estimates include *structural models*, *economic-demographic models*, *dual system estimation*, *microsimulation models*, *neural networks*, the *grouped answer method*, *social network analysis*, and various methods related to *spatial demography* (Swanson and Tayman, 2012: 219–242). A *local census*, which provides a direct count of small area populations independent of any other data source, can also be used, though it is expensive and its accuracy dependent on the response rate achieved (Simpson, Diamond, Tonkin and Tye, 1996; Lunn, Simpson, Diamond and Middleton, 1998).

In addition to individual techniques of population estimation, one can also estimate the population by the composite method and by combining estimates. The *composite method*, which was developed by Bogue and Duncan (1959), is a "portfolio" of separate estimation methods that are tailored to particular segments of the population such as age groups (Bryan, 2004b: 550–551). For example, one could apply the censal-ratio method using births to develop an estimate for the population aged 0–4, the component method using school enrollment data for the population aged 5–17, the censal-ratio method using births and deaths for the population aged 18–64, and administrative records using Medicare data for the population aged 65 and older. The estimates for each age group are then summed up to yield an estimate of total population. Many combinations of methods and data are available, though there exists little empirical guidance in terms of which to choose (Siegel, 2002: 416–417). Another approach is to *combine estimates* made with two or more different estimation techniques. In its most simple way, this involves averaging or weighting the estimates directly; alternatively, one can join estimates from sample surveys and estimates based on demographic analysis as independent variables in a regression model. There is evidence suggesting that combining estimates made with different methods, data, and assumptions improves accuracy and reduces bias (Hoque, 2012; Simpson, Diamond, Middleton, Lunn and Cossey, 1998).

In recent years, much research has been devoted to small area population estimation methods that utilize remote sensing and GIS technologies (Wu, Qiu and Wang, 2005; Wang and Wu, 2010). Dasymetric mapping – the redistribution of areal enumeration data using ancillary information for the display of statistical surface data – once confined to visualizing population distribution, is also increasingly being used for estimating the population of small areas (Mennis, 2009; Petrov, 2012). These methods are attractive since they are not tied to the administrative boundaries associated with census or survey data and allow population estimates to be made for user defined areas. They also show potential for areas where census data and administrative records are unavailable, unreliable, or out of date, as is the case in many countries in the developing world (Hardin and Shumway, 2008; Sutton, Roberts, Elvidge and Baugh, 2001). While promising, there currently exists a disconnect between academic research and the actual adoption and use of remotely sensed technologies for population estimation purposes, with lack of expertise, technical, application, and financial issues preventing more widespread adoption by practitioners (Hoalst-Pullen and Patterson, 2011).

Methods for Small Area Population Projections

Many of the methods used for producing population estimates can also be applied when preparing population projections and the sources of data and factors to be considered when choosing among methods are similar as well. There are important differences, though, since estimates are concerned with giving an accurate population count for a past or present date, whereas projections require that assumptions about future population changes be made. Projections involve more uncertainty, and projection methods vary widely in how they attempt to model it. As was the case with estimates, there are various ways to classify projection methods. For small areas, Smith, Tayman and Swanson (2013) propose four types: trend extrapolation, cohort-component, structural, and microsimulation models.

Trend extrapolation models determine future population values solely by their historical values; they range in complexity from simple methods such as linear, geometric, and exponential growth to complex polynomial, logistic curve fitting, and ARIMA time series models. *Ratio methods* – in which the population or population change of a small area is expressed as a proportion of population or population change of a larger area in which the smaller area is located – is another type of extrapolation that is often used (Smith, Tayman and Swanson, 2013: 185–213). Recent research developments involve the application of ratio methods in combination with GIS techniques to produce small area projections based on grids

(Hachadoorian, Gaffin and Engelman 2011). Trend extrapolation projection models have a number of shortcomings: they do not account for differences in demographic composition or for differences in the components of growth, provide little or no information on the projected demographic characteristics of the population, have no theoretical content and generally cannot be related to theories of population change. On the other hand, the small data requirements, ease of application and explanation, timeliness, and low cost of simple trend and ratio models make them an attractive option particularly for small areas where data availability and reliability issues may preclude the application of more complex projection models (George, Smith, Swanson and Tayman, 2004: 571). Furthermore, simple trend models often provide reasonably accurate projections for short and even long horizons, and the empirical evidence suggests that more complex and sophisticated models do not offer more accurate projections, at least for total population (Armstrong, 1984; Chi, 2009; Smith, 1997).

C*ohort-component models* are the most commonly used population projection method. Going back to the work of Cannan (1895), the cohort-component method accounts separately for the three components of population change: births, deaths, and migration. Most models subdivide the population into age and sex groups, and each demographic subgroup is projected separately; further subdivisions by race, ethnicity, or other demographic characteristics are possible. While assumptions about the future need to be made for all three components of population change in a cohort-component model, migration tends to be the most significant, most volatile, and hardest to predict for small areas. Approaches for projecting migration include methods which use only base period migration data, methods which require some additional qualitative or quantitative information, and methods based on quantitative projections of independent variables (Wilson, 2011: 33–36). Cohort-component models vary further in whether they apply gross or net migration data. Gross migration models range from complex migration pool and multi-regional models that are primarily useful for larger areas to simpler bi-regional models that are easier to apply for small areas. Net migration models can follow a top-down or bottom-up approach; a simplified version proposed by Hamilton and Perry (1962) that treats mortality and migration as a single unit is also sometimes used (Smith, Tayman and Swanson, 2013: 155–183). The cohort-component method is a very popular projection method because it can incorporate many different data sources, assumptions, and application techniques. It is well suited to provide projections of demographic characteristics in addition to projections of total population. However, the method is also very data intensive and relatively expensive to apply, and the potential lack of complete and reliable data can be a serious obstacle when projecting the population for small areas (Smith and Morrison, 2005).

Structural models form the third major group of projection methods. In a structural model, population change is related to changes in one or more explanatory variables, and the focus is generally on modeling migration. There are two major types of structural models: *economic-demographic* and *urban system models*. Examples of the former include econometric models, models that balance labor supply and demand, models based on population/employment ratios, and regional economic models. Economic-demographic models generally focus on economic variables to project migration, though some include amenities as well; most models are complex, resource intensive, expensive to develop and apply, and primarily used for relatively large areas such as counties, metropolitan areas, states, and nations (Smith, Tayman and Swanson, 2013: 216–228). For smaller areas, urban system models are more commonly applied; these typically include land use and transportation characteristics in addition to economic variables and rely heavily on GIS techniques. Like most economic-demographic models, urban systems models tend to be very resource intensive and often require a substantial degree of technical expertise (Smith, Tayman and Swanson, 2013: 228–237). While not necessarily more accurate than other types of projection methods, structural models are useful in that they can address a wide range of theoretical, planning, and policy questions. They can make important contributions to the planning and decision-making process and are particularly well suited for simulation and scenario purposes (George, Smith, Swanson and Tayman, 2004: 586).

While most population projections for small areas are made with trend extrapolation, cohort-component, or structural methods, a number of additional approaches are worth mentioning. The *housing unit method*, though more commonly applied to prepare population estimates, also shows potential for small area population projections (Foss, 2002). The method takes into account local housing supply and, either explicitly or implicitly, residential land availability, and can be used to constrain cohort-component projections. It promises to be particularly useful for growing urban areas and areas earmarked for residential development, though projecting its various components is challenging (Wilson, 2011: 20–22). *Microsimulation models* differ from traditional demographic modeling as follows: they use a sample rather than the total population; they work on the level of the individual or the household rather than with grouped data; and they rely on repeated random experiments to derive a projection (van Imhoff and Post, 1998). By modeling individual behavior, spatial microsimulation models avoid aggregation bias, are internally consistent, provide very detailed projection outputs, and allow for scenario and what-if analyses; on the downside, the models tend to be extremely complex, require extensive data and staff resources, and there have been few attempts to validate the assumptions and outcomes of these models (Birkin and Clarke, 2011; Wilson, 2011: 37–38). Another avenue of recent research involves the incorporation of spatial effects into the projection model. Originating in the geographic concept of spatial diffusion – the spread of a particular phenomenon over space and time – *spatial diffusion* and *spatio-temporal projection models* incorporate population spillovers and legacy effects arising from neighboring growth rates and characteristics into the model (Thrall, Sidman, Thrall and Fik, 2001; Chi and Voss, 2011). While these methods address a major shortcoming of most projection models used for small areas – which treat each unit of geography as an independent, stand-alone entity rather than as an entity surrounded by other areas with which they interact – they have not been found to outperform forecasts made with simple extrapolation methods (Chi and Voss, 2011; Chi, Zhou and Voss, 2011).

As was true with estimates, *combining projections* using simple averages, weighted averages, trimmed means, and composite approaches, is another option. Combined projections are ideally derived from methods that differ substantially from one another and that draw from different sources of information; combining has been found to improve forecast accuracy under many conditions (Ahlburg, 1995; Armstrong, 2001; Armstrong, 2006; Clemen, 1989; Rayer, 2008). Finally, in addition to projections of total population and by demographic characteristics, many types of planning, budgeting, and analysis also require projections of households, school enrollment, employment, health, poverty, or other population-related characteristics. While structural models are often applied for these purposes, the *participation-ratio* and *cohort-*

progression methods are worthwhile to consider. In the former, socioeconomic characteristics are related to demographic characteristics through the use of ratios; in the latter, projections are developed by "surviving" people with particular socioeconomic characteristics (George, Smith, Swanson and Tayman, 2004: 586–590).

Among the most important advances in recent years has been the development of *probabilistic projection methods* (Wilson and Rees, 2005). Probabilistic methods acknowledge that forecasting the population entails a significant amount of uncertainty, especially for longer time periods, for places with small or rapidly changing populations, for certain age groups, and for the demographic components of population change. Conventional deterministic projections provide a single population number, or high and low variants that do not attach probability to the high-low ranges. Probabilistic methods, in contrast, assess and communicate uncertainty by providing forecasts that come with probability distributions or that are fully probabilistic, i.e. generated by probabilistic population renewal (Lee, 1998). Predictive intervals are commonly made with three distinct approaches: (1) time-series analyses; (2) expert-based probabilistic projections; and (3) ex post analysis, i.e. the use of past forecast errors (Bongaarts and Bulatao, 2000: 200–204). Most of the research on probabilistic projection methods has focused on large geographic areas such as nations, states, and regions. Yet a probabilistic approach is arguably even more important for small areas, given the greater uncertainty of future population changes at the subnational level (Tayman, 2011; Wilson and Rees, 2005). Cameron and Poot (2011) discuss some of the challenges of providing probabilistic projections for small areas and offer directions for further methodological development.

From Administrative Units to User Defined Areas?

In the first edition of this chapter, Alho (2001: 3484) opined that "with the development of the geographic information system (GIS), small area estimation and forecasting will no longer be limited by administrative boundaries." Arguing that administrative boundaries are frequently not well suited to many types of social and medical research, and that domains of interest may intersect several areas or be too small for their effect to be discernible in small area data, the introduction of GIS techniques was viewed as a major step forward for small area estimation and forecasting (Alho, 2001: 3485–3486). Smith, Tayman and Swanson (2001: 365–367) also predicted a bright future for GIS in small area analysis, but argued that while GIS techniques may be particularly useful when combined with remote sensing to develop small-area population and housing estimates, for projections the main value of GIS would likely be in developing databases rather than in constructing projections per se. Now, more than a decade later, one can ask: to what extent have small area estimates and projections incorporated spatial techniques, and how much progress has been made to overcome the limitations of estimating and projecting populations based on administrative units?

As discussed above, in recent years much research has focused on population estimation using remote sensing and GIS techniques. Not only have these techniques created new and alternative means of acquiring population estimates for small areas previously estimated with traditional methods, they also allow estimates to be made for user defined areas that were hard to come by in the past. On the other hand, the literature on projections made using GIS techniques remains sparse. Projections require assumptions about the future, and while GIS and spatial techniques can add to our understanding of past changes and present population distributions, they do not make forecasting them easier. It seems likely that for the foreseeable future GIS and spatial techniques will remain most useful as tools for distributing populations projected with established demographic models to areas smaller than, or different from, those they were originally made for, rather than providing an alternative projection approach.

Most small area estimates and projections are still made for administrative units, and a major shift towards user defined areas is unlikely to happen soon. Most of the required input data are still collected for administrative units, and methods that allow for customized bottom-up aggregations – such as microsimulation models – or methods that transcend administrative boundaries – such as remote sensing and GIS techniques – have not yet replaced estimates and projections made with established demographic models. Administrative units can be inconvenient to work with, but they often represent meaningful entities for which knowledge about past and present population changes exists. While small area population estimates and projections will always contain an element of uncertainty, this professional knowledge, when combined with a careful selection of data and methods, still provides the best foundation for satisfactory estimation and forecasting outcomes.

See also:

Censuses: Current Approaches and Methods; Demographic Models; Demographic Techniques: Indirect Estimation; Geographic Information Systems; Microsimulation in Demographic Research; Population Forecasts; Population Geography; Remote Sensing; Urban Planning: Methods and Technologies.

References

Ahlburg, Dennis A. (1995). Simple versus complex models: evaluation, accuracy, and combining. *Mathematical Population Studies* **5**, 281–290.

Alho, Juha M. (2001). Demographic techniques: small-area estimates and projections. In Smelser, Neil J. and Baltes, Paul B. (eds.) *International Encyclopedia of the Social and Behavioral Sciences*, pp. 3483–3486. Amsterdam & New York: Elsevier.

Armstrong, J. Scott (1984). Forecasting by extrapolation: conclusions from 25 years of research. *Interfaces* **14**(6), 52–66.

Armstrong, J. Scott (2001). Combining forecasts. In Armstrong, J. Scott (ed.) *Principles of forecasting: a handbook for researchers and practitioners*, pp. 417–439. Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. Scott (2006). Findings from evidence-based forecasting: methods for reducing forecast error. *International Journal of Forecasting* **22**, 583–598.

Baker, Jack, Ruan, Xiaomin, Alcantara, Adelamar, Jones, Troy, Watkins, Kendra, McDaniel, Michael, Frey, Margaret, Crouse, Nathan, Rajbhandari, Ruji, Morehouse, Jana, Sanchez, Jeremy, Inglis, Mike, Baros, Shirley, Penman, Shawn, Morrison, Susan, Budge, Tom and Stallcup, Wes (2008). Density-dependence in urban housing unit growth: an evaluation of the Pearl-Reed model for predicting housing unit stock at the census tract level. *Journal of Economic and Social Measurement* **33**, 155–163.

Birkin, Mark and Clarke, Martin (2011). Spatial microsimulation models: a review and a glimpse into the future. In Stillwell, John and Clarke, Martin (eds.) *Population dynamics and* *projection methods, Understanding population trends and processes, Volume 4* (1st edn.), pp. 193–208. Dordrecht, The Netherlands: Springer.

Bogue, Donald J. (1950). A technique for making extensive population estimates. *Journal of the American Statistical Association* **45**(250), 149–163.

Bogue, Donald J. and Duncan, Beverly (1959). A composite method for estimating postcensal population of small areas by age, sex, and color. In *Vital Statistics - Special Reports*. Washington, D.C.: National Office of Vital Statistics.

Bongaarts, John and Bulatao, Rodolfo A. (eds.) (2000). *Beyond six billion: forecasting the world's population*. Washington, DC: National Academy Press.

Bryan, Thomas (2004a). Basic sources of statistics. In Siegel, Jacob S. and Swanson, David A. *The methods and materials of demography* (2nd edn.), pp. 9–41. San Diego, California: Elsevier Academic Press.

Bryan, Thomas (2004b). Population estimates. In Siegel, Jacob S. and Swanson, David A. *The methods and materials of demography* (2nd edn.), pp. 523–560. San Diego, California: Elsevier Academic Press.

Cameron, Michael P. and Poot, Jacques (2011). Lessons from stochastic small-area population projections: the case of Waikato subregions in New Zealand. *Journal of Population Research* **18**, 245–265.

Cannan, Edwin (1895). The probability of as cessation of the growth of population in England and Wales during the next century. *The Economic Journal* **5**(20), 505–515.

Chi, Guangqing (2009). Can knowledge improve population forecasts at subcounty levels? *Demography* **46**(2), 405–427.

Chi, Guangqing and Voss, Paul R. (2011). Small-area population forecasting: borrowing strength across space and time. *Population, Space and Place* **17**, 505–520.

Chi, Guangqing, Zhou, Xuan and Voss, Paul R. (2011). Small-area population forecasting in an urban setting: a spatial regression approach. *Journal of Population Research* **28**, 185–201.

Clemen, Robert T. (1989). Combining forecasts: a review and annotated bibliography. *International Journal of Forecasting* **5**, 559–583.

Coulton, Claudia J. (2008). *Catalog of administrative data sources for neighborhood indicators*. Washington, D.C.: The Urban Institute.

Deng, Chengbin, Wu, Changshan and Wang, Le (2010). Improving the housing-unit method for small-area population estimation using remote-sensing and GIS information. *International Journal of Remote Sensing* **31**(21), 5673–5688.

Foss, Wayne (2002). Small area population forecasting. *The Appraisal Journal* **70**(2), 163–172.

George, M. V., Smith, Stanley K., Swanson, David A. and Tayman, Jeff (2004). Population projections. In Siegel, Jacob S. and Swanson, David A. *The methods and materials of demography* (2nd edn.), pp. 561–601. San Diego, California: Elsevier Academic Press.

Ghosh, M. and Rao, J. N. K. (1994). Small-area estimation – an appraisal. *Statistical Science* **9**(1), 55–76.

Hachadoorian, Lee, Gaffin, Stuart R. and Engelman, Robert (2011). Projecting a gridded population of the world using ratio methods of trend extrapolation. In Cincotta, Richard P. and Gorenflo, Larry J. (eds.) *Human population: its influences on biological diversity*, pp. 13–25. Ecological Studies, Volume 214. Berlin, Germany: Springer.

Hamilton, C. Horace and Perry, Josef (1962). A short method for projecting population by age from one decennial census to another. *Social Forces* **41**(2), 163–170.

Hardin, Perry J. and Shumway, J. Matthew (2008). Intraurban population estimation with remotely sensed imagery: a developing world perspective. *Geography Compass* **2**(3), 744–770.

Hoalst-Pullen, Nancy and Patterson, Mark W. (2011). Applications and trends of remote sensing in professional urban planning. *Geography Compass* **5**(5), 249–261.

Hoque, Nazrul (2012). Evaluation of small area population estimates produced by housing unit, ratio-correlation, and component method II compared to 2000 census counts. *Canadian Studies in Population* **39**(1-2), 91–108.

Lee, Ronald D. (1998) Probabilistic approaches to population forecasting. *Population and Development Review* **24**, 156–190.

Lunn, David J., Simpson, Stephen N., Diamond, Ian and Middleton, Liz (1998). The accuracy of age-specific population estimates for small areas in Britain. *Population Studies* **52**(3), 327–344.

Martin, David (2006). Last of the censuses? The future of small area population data. *Transactions of the Institute of British Geographers* **31**(1), 6–18.

Mennis, Jeremy (2009). Dasymetric mapping for estimating population in small areas. *Geography Compass* **3**(2), 727–745.

Murdock, Steve H., Cline, Michael and Zey, Mary (2012). Challenges in the analysis of rural populations in the United States. In Kulcsar, Laszlo J. and Curtis, Katherine J. (eds.)

International handbook of rural demography, pp. 7–15. Dordrecht, The Netherlands: Springer.

Murdock, Steve H., Hwang, S. and Hamm, R. R. (1995). Component methods. In Rives, N. W, Serow, W. J., Lee, A. S., Goldsmith, H. F. and Voss, P. R. (eds.) *Basic methods for preparing small-area estimates*, pp. 10–53. Madison, Wisconsin: Applied Population Laboratory, University of Wisconsin.

National Research Council – Committee on the Effective Use of Data, Methodologies, and Technologies to Estimate Subnational Populations at Risk (2007). *Tools and methods for estimating populations at risk from natural disasters and complex humanitarian crises*. Washington, DC: National Academies Press.

O'Muircheartaigh, Colm (1997). Measurement error in surveys: a historical perspective. In Lyberg, Lars, Biemer, Paul, Collins, Martin, de Leeuw, Edith, Dippo, Cathryn, Schwarz, Norbert and Trewin, Dennis (eds.) *Survey measurement and process quality*, pp. 1–25. New York: Wiley.

Petrov, Andrey (2012). One hundred years of dasymetric mapping: back to the future. *The Cartographic Journal* **49**(3), 256–264.

Pfeffermann, Danny (2002). Small area estimation – new developments and directions. *International Statistical Review* **70**(1), 125–143.

Popoff, Carole and Judson, D. H. (2004). Some methods of estimation for statistically underdeveloped areas. In Siegel, Jacob S. and Swanson, David A. The methods and materials of demography (2nd edn.), pp. 603–641. San Diego, California: Elsevier Academic Press.

Rao, J. N. K. (2003). *Small area estimation*. Hoboken, New Jersey: Wiley.

Rayer, Stefan (2008). Population forecast errors: a primer for planners. *Journal of Planning Education and Research* **27**, 417–430.

Raymondo, James C. (1992). *Population estimation and projection: Methods for marketing, demographic, and planning personnel*. New York: Quorum Books.

Rives, N. W., Serow, W. J., Lee, A. S. and Goldsmith, H. F. (1989). *Small area analysis: estimating total population*. Rockville, Maryland: U.S. Department of Health and Human Services.

Robson, Colin (2011). *Real world research* (3rd edn.). Chichester, United Kingdom: Wiley.

Schmitt, Robert C. and Crosetti, Albert H. (1954). Accuracy of the ratio-correlation method for estimating post-censal population. *Land Economics* **30**(3), 279–281.

Siegel, Jacob S. (2002). *Applied demography: applications to business, government, law, and public policy*. San Diego, California: Academic Press.

Simpson, Stephen, Diamond, Ian, Tonkin, Pete and Tye, Rachel (1996). Updating small area population estimates in England and Wales. *Journal of the Royal Statistical Society, Series A* **159**, 235–247.

Simpson, Stephen, Diamond, Ian, Middleton, Elizabeth, Lunn, David and Cossey, Rachel (1998). How can a national set of small area population estimates gain acceptance? Lessons from the Estimating with Confidence project. In Arnold, Richard, Elliott, Paul, Wakefield, Jon and Quinn, Mike (eds.) *Population counts in small areas: implications for studies of environment and health*. London, England: The Stationary Office.

Simpson, Stephen, Middleton, Liz, Diamond, Ian and Lunn, David (1997). Small-area population estimates: a review of methods used in Britain in the 1990s. *International Journal of Population Geography* **3**(3), 265–280.

Smith, Stanley K. (1986). A review and evaluation of the housing unit method of population estimation. *Journal of the American Statistical Association* **81**(394), 287–296.

Smith, Stanley K. (1997). Further thoughts on simplicity and complexity in population projection models. *International Journal of Forecasting* **13**(4), 557–565.

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

Smith, Stanley K. and Morrison, Peter A. (2005). Small-area and business demography. In Poston, Dudley and Micklin, Michael (eds.) *Handbook of population*, pp. 761–785. New York: Kluwer Academic/Plenum Publishers.

Smith, Stanley K., Tayman, Jeff and Swanson, David A. (2013). *A practitioner's guide to state and local population projections.* Dordrecht: Springer.

Smith, Stanley K., Tayman, Jeff and Swanson, David A. (2001). *State and local population projections: methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Starsinic, D. E., Lee, A. S., Goldsmith, H. F. and Sparr, M. A. (1995). The Census Bureau's administrative records method. In Rives, N. W, Serow, W. J., Lee, A. S., Goldsmith, H. F. and Voss, P. R. (eds.) *Basic methods for preparing small-area estimates*, pp. 54–69. Madison, Wisconsin: Applied Population Laboratory, University of Wisconsin.

Sutton, P., Roberts, D, Elvidge, C. and Baugh, K (2001). Census from heaven: an estimate of the global human population using night-time satellite imagery**.** *International Journal of Remote Sensing* **22**(16), 3061–3076.

Swanson, David A. (2004). Advancing methodological knowledge within state and local demography: a case study. *Population Research and Policy Review* **23**(4), 379–398.

Swanson, David A. and Beck, Donald M. (1994). A new short-term county population projection method. *Journal of Economic and Social Measurement* **20**, 25–50.

Swanson, David A. and Hough Jr., George C. (2012). An evaluation of persons per household (PPH) estimates generated by the American Community Survey: a demographic perspective. *Population Research and Policy Review* **31**(2), 235–266.

Swanson, David A. and McKibben, Jerome N. (2010). New directions in the development of population estimates in the United States? *Population Research and Policy Review* **29**, 797–818.

Swanson, David A. and Tayman, Jeff (2012). *Subnational population estimates*. Dordrecht, The Netherlands: Springer.

Tayman, Jeff (2011). Assessing uncertainty in small area forecasts: state of the practice and implementation strategy. *Population Research and Policy Review* **30**, 781–800.

Thrall, Grant Ian, Sidman, Charles F., Thrall, Susan Elshaw and Fik, Timothy J. (2001). The Cascade GIS diffusion model for measuring housing absorption by small area with a case study of St. Lucie County, Florida. *The Journal of Real Estate Research* **8**(3), 401–420.

van Imhoff, Evert and Post, Wendy (1998). Microsimulation methods for population projection. *Population: An English Selection* **10**(10), 97–138.

Wang, Le and Wu, Changshan (2010). Preface: population estimation using remote sensing and GIS technologies. *International Journal of Remote Sensing* **31**(21), 5569–5570.

Wilson, Tom (2011). A review of sub-regional population projection methods. *Queensland Centre for Population Research, The University of Queensland, Australia*.

Wilson, Tom, and Rees, Phil (2005). Recent developments in population projection methodology: a review. *Population, Space and Place* **11**, 337–360.

Wu, Shou-sheng, Qiu, Xiaomin and Wang, Le (2005). Population estimation methods in GIS and remote sensing: a review. *GIScience and Remote Sensing* **42**(1), 58–74.