Spatial Demography (2023) 11:8
https://doi.org/10.1007/s40980-023-00119-3

An Example of Combining Expert Judgment and Small Area Projection Methods: Forecasting for Water District Needs

David Swanson[1,2,3] · Tom Bryan[4] · Mark Hattendorf[5] · Kelly Comstock[6] · Lauren Starosta[6] · Robert Schmidt[5]

Accepted: 27 July 2023

Abstract
This case study shows how GIS and Expert Judgment can be used to develop small area population forecasts in the United States. It starts by organizing 2000 and 2010 block group population data into the 2020 block group geography and then examines 2020 indicators used to evaluate the effect of Differential Privacy on the 2020 population data. Preliminary population projections to 2050 are then generated by averaging the results of three standard small area projection methods. Using local expert judgment, GIS overlay maps and satellite imagery in a virtual environment, the 301 block groups of Greenville County, South Carolina were classified into seven categories of future population change. These categories were then applied to the preliminary projections to generate informed forecasts. Following this step, the sums of the BG results were then compared, respectively, to independently generated county population forecasts for 2030, 2040, and 2050. At this point, 25 BGs were selected for additional review, which resulted in a final set of forecasts. We find that the increase of 152,840 people in the year 2050 spread over all of the 301 census block groups in going from the preliminary projections (675,626) to the final informed forecasts (828,467) is largely generated by these same 25 BGs, which expert judgment determined were currently poised to "take off" in terms of population growth. Having this much change generated by such a small number of BGs is consistent with findings elsewhere.

Keywords GIS · Differential privacy · Satellite imagery · Model-based theory · Virtual environment

1 Introduction

Forecasting is always at the forefront of decision-making and planning (Petropoulos et al., 2022) and in this paper, we use the case study approach (Swanson & Morrison, 2010) to provide an example of population forecasting that we believe is at the

Extended author information available on the last page of the article

Published online: 10 August 2023

leading edge of this forefront because it combines expert opinion and algorithmic procedures, a suggestion found in Zellner et al. (2021). Before turning to a description of the study and our results, however, we first describe our major objectives, which is followed by a discussion of key concepts and terminology in regard to population forecasting. After this discussion we provide a conceptual overview of the strengths and weaknesses of population projection methods. The paper then proceeds with a description of the data and methods, followed by a description of the results. The paper concludes with a discussion of the results in which some time is spent on the current and future state of small area population data, to include forecasting and estimation because of the effect on the reliability of small area data caused by the decision of the U.S. Census Bureau to apply "differential privacy" to the 2020 census as its new disclosure avoidance system (Hauer & Santos-Lozada, 2021; Hotz & Salvo, 2022; Ruggles & Van Riper, 2022; Sullivan, 2020, pp. 69–83; Swanson & Cossman, 2021; Swanson et al., 2022; Winkler et al., 2021).

1.1 Major Objectives

The major objectives we have in this paper are fourfold. First, we present the case that all population forecasts require the exercise of judgment and theory and that the former needs to be based on knowledge and experience while the algorithmic method(s) used in calculating population projections should be grounded in demographic theory.

Second, we argue that a given forecasting method's strengths and weaknesses largely stem from four sources: (1) Its correspondence to the dynamics by which a population moves forward in time; (2) the information available relevant to these dynamics; (3) the time and resources available to assemble the information required by one or more theoretically-grounded algorithmic method(s) to generate one or more projections and then apply knowledge and experience to the projection(s) in order to generate a forecast; and (4) the information needed from the forecast.

Third, by tying together the first and second objectives we present the argument that as experience, objectivity and knowledge increase, the accuracy of the forecast is likely to increase when it is based on one or more theoretically-grounded algorithmic method(s).

Fourth, we describe in this paper a process that illustrates how small area population forecast accuracy can be improved by linking these first and second objectives and conclude by discussing the results of our linkage in this case study.

1.2 Conceptual Background and Terminology

In regard to the future of a given population, expert opinion needs to be based on knowledge and experience (Dalkey, 1968; Hogarth & Makridakis, 1981; Kahneman, 2011; Theocharis & Harvey, 2019) and algorithmic procedures (i.e., projection methods) should be grounded in demographic theory (Burch, 2018; Petropoulos et al., 2022). Before proceeding with a discussion of the strengths and weaknesses of these two approaches, we first define: (1) a population projection as the outcome of a set of conditional statements made by a person, group, or agency about the future in that they show what the population in question would be *if* particular assumptions were to hold true, but make no prediction as to whether those assumptions actually *will* hold true; and (2) a population forecast as the projection a person, group, or agency believes is most likely to provide an accurate prediction of the future population (Smith et al., (2002, p. 3), which means that all population forecasts are explicitly judgmental. Because a population projection can be generated by a subjective process, a quantitative process, or a hybrid of the two, a population forecast also is generated by them. In accordance with this view, we interpret "expert opinion" as a subjective process, "algorithmic procedures" as quantitative processes, and a combination of the two as a "hybrid" process.

1.3 "Strengths and Weaknesses" of Subjective and Quantitative Approaches

Because no projection method, subjective or quantitative, yields a forecast unless judgment is applied, a given method's strengths and weaknesses largely stem from four sources: (1) Its correspondence to the dynamics by which a population moves forward in time; (2) the information available relevant to these dynamics; (3) the time and resources available to assemble relevant information and generate a forecast; and (4) the information needed from the forecast. In terms of population dynamics, it is the fundamental population equation that applies, namely that the population at a given point in time, P_{t+k}, is equal to the population at an earlier point in time, P_t, to which is added the births and in-migrants that occur between time t and time $t+k$ and to which is subtracted the deaths and out-migrants that occur during this same time period (Baker et al., 2017, pp. 251–252): $P_{t+k} = P_t + \text{Births} + \text{In Migration—Deaths} - \text{Out Migration}$. This equation not only applies to the population as a whole but also to age groups, age-gender groups, age-gender-race groups, and so on in regard to a range of demographic characteristics (Baker et al., 2017, pp. 191–208). This makes the fundamental equation the cornerstone of demographic theory. Because it is the basis for the cohort-component method of population projection population, it also places this method in the foundation of demographic theory, which as its name suggests also incorporates the cohort perspective on population change, which itself is fundamental both to population dynamics and demographic theory. Moreover, because the "Cohort Change Ratio Method," a population projection method we employ in this case study, is algebraically equivalent to the fundamental population equation, and, as its name suggests, also deals with cohort change, this method can be viewed as fundamental to population dynamics and demographic theory (Baker et al., 2017; Swanson et al., 2016).

Barring unforeseeable catastrophes and other "Black Swan" events that have very low probabilities of occurring (Taleb, 2010), the closer one comes to having accurate data concerning P_t, and the birth, deaths, in-migrants, and out-migrants corresponding to P_{t+k}, the stronger the projection method will likely be; as the accuracy of P_t and the direct correspondence of births, deaths, im-migrants, and out-migrants to P_t lessens, the weaker it will likely be. If, as an example, P_t is the total population of Canada in 2020, one can assemble birth, death, and migration directly corresponding to it relatively quickly and use the Cohort Component Method (Smith et al., 2002, 2013) to project it using an "off-the-shelf" template within a relatively short period of time at low cost. If, however, one needs to forecast the population of Hopi tribal members living on the Hopi Reservation in Arizona by age and sex, one may have to start with a less-than-accurate census count of this population (Swanson, 2021) by age and gender and apply birth, death, and migration data to it from one or more analogue populations (Swanson, 2022). Because it is virtually impossible to assemble birth, death, and migration data that directly correspond to it in any reasonable length of time even if one has substantial resources available for this assembly, one will likely be forced to select one or more analogues (Swanson, 2022).

Turning back to the example of Canada, if one is under severe time and resource constraints, the "Cohort Change Ratio" method could be used (Baker et al., 2017), which is likely to yield neither a total population projection nor, if desired, a projection by age and sex much different from that generated by the Cohort-Component Method, respectively (Baker et al., 2017). Similar observations apply to extrapolative methods, including simple ones such as those that are linear or exponential and those that are more complex, such as ARIMA (Auto-Regressive Integrated Moving Average) and other forms of time-series models; they also apply to structural models (Smith et al., 2002, 2013). In regard to subjective and quantitative methods, Zellner et al. (2021) find that neither is universally superior and in regard to simple and complex methods, Green and Armstrong (2015) find that complex methods provide no more accuracy than do simple methods.

The strengths and weaknesses in a given forecast will be based on those just discussed in the underlying projection method(s) and, also, on the experience, objectivity and knowledge of the person(s) judging that a given projection is more likely to represent the future of the population in question than is another projection As experience, objectivity and knowledge increase, the accuracy of the forecast is likely to increase; as they decrease, so is the accuracy of the forecast. Whether it is the judgment of a single individual (rule-based or otherwise), or that of a group (in the form of rule-based approaches such as Delphi or a focus group, or otherwise), one needs to consider, as was the case with projection methods, the issue of "utility"—the likely level of accuracy vs. the time and resources required to attain this level (Swanson & Tayman, 1995, 1996). It is also the case that a "less accurate" forecast may yield a forecast that is sufficient for a given need than the level required for another need (Swanson & Tayman, 1996; Swanson et al., 1997, 1998).

In regard to a new development, machine learning (aka artificial intelligence), a recent study by Baker et al. (2023) suggests that this approach may improve accuracy over traditional projection methods with a reasonable level of utility, but if such a projection is to be considered a forecast, it is inescapable that human judgment will be employed, which means we are again looking at the experience, objectivity, and knowledge of the person(s) making such a judgment.

2 Background

Turning now to our case study, "Greenville Water" (https://www.greenvillewater.com), a public agency, located in Greenville, South Carolina, owns all of the 26,000 acres of two watersheds in which are three protected mountain reservoirs: (1) Table Rock – at the head of the South Saluda River; (2) Poinsett – at the head of the North Saluda River; and (3) Lake Keowee – in western Pickens County. Greenville Water signed a Conservation Easement with The Nature Conservancy so that the land use remains the way it is, with no construction or development now or in the future. Along with the three reservoirs, Greenville Water operates treatment plants and a testing facility aimed at delivering high-quality water to over 350,000 Greenville residents. Figure 1 shows Greenville County, SC, where Greenville Water is located.

Overall, population growth has been substantial and dynamic, being driven by some areas with very recent and substantial development. Other areas are expected to follow suit, including Greenville City, which is undergoing redevelopment; However, there are some areas in modest decline. In this environment, not only knowing how parts of the county have changed in the past, but which parts of the county are expected to experience population increase or decline in the future is critically important for accurate, efficient planning.

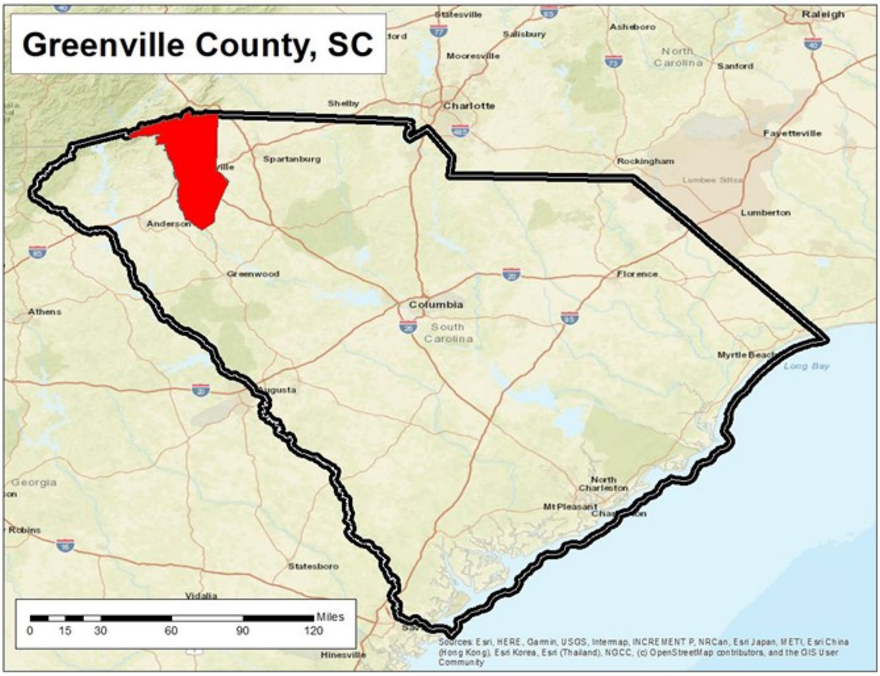

Fig. 1 Greenville County, South Carolina

With the population of its service area expected to increase overall, Greenville Water embarked on planning aimed at increasing its infrastructure to service the demand out to the year 2050. It is expected to increase because of the growth observed from 2000 to 2010 (from 379,635 to 451,283, a 19% increase) and again from 2010 to 2020 (from 451,283 to 525,534, a 16.5% increase). As noted in the Introduction, forecasts are at the forefront of decision-making and planning and population forecasts are a core component of water use planning (Miro et al., 2018). Previous planning processes and expansion had generated population forecasts and there were forecasts done by other entities that could be used (See Table 1) and accompanying discussion. However, with the indications that population growth in the Greenville Water service area would be substantial, professional demographers were retained to develop the population forecasts, which for planning purposes, needed to be done by block group (BG), which according to the 2020 census numbered 301 for Greenville County as a whole.

Table 1 Comparison of selected Greenville county projections

Source	2020[a]	2030	2040	2050
Provisional forecast informed by expert judgment	525,534	619,752	731,137	837,898
Preliminary naïve projection	525,534	602,792	697,161	787,245
ARIMA[b]	525,534	568,555	670,828	859,734
Hamilton-Perry method[c]	525,534	607,466	682,821	759,090
State of SC4	525,534	616,105	706,862	812,599
2009 Master plan[d]	525,534	549,010	N/A	N/A
2014 Keowee-Taxaway[d]	525,534	600,000	675,000	N/A
2014 ACOG[d]	525,534	596,000	670,000	N/A

[a]Regardless of what a projection completed prior to 2020 showed for 2020, the US. Census Bureau's 2020 Census Total for Greenville County is displayed in the table

[b]ARIMA stands for "Auto Regressive Integrated Moving Average" (Smith et al., 2013, 199–209). This is a time series model that using decennial census numbers and intercensal annual estimates generated by the U.S. Census Bureau for the period from 1970 to 2020 as input to generate annual numbers from 2021 to 2050

[c]The Hamilton-Perry Method is a variation of the Cohort Component Method of Population projection that requires only age data for the population in question at two successive census counts or estimates (Baker, Swanson, Tayman, and Tedrow, 2017). Here the data for the 2010 census are used in conjunction with a 2015 estimate done by the U.S. Census Bureau are used as input with the projection launched from 2015

[d]These are taken from Fig. 4.2 in the 2016 Master Plan for the Greenville Water District (2016) and are approximate numbers. None of them go beyond 2040. However, because we have the actual projections done by the South Carolina (State of SC) from a separate source, we extended this projection to 2050 using the rate of change found between 2030 and 2040

2.1 Data and Methods

Because population growth is a critical element in future water demand (Butler & Memon, 2005, p. xiii), population forecasts are key to developing a forecast of water demand. Not surprisingly, different population projection methods have been used in the development of these forecasts. They include the cohort-component method, structural models and time-series-based trend extrapolation methods (Miro et al., 2018, p. 11; Smith et al., 2002: Texas Water Development Board, 1997, 2021). Typically, a water demand forecast contains two elements, per-customer use and the number of customers (Miro et al., 2018, 4). In the case of Greenville Water, the customers are people, which makes the number of customers equivalent to the total population. The per-capita use is typically determined by historical demand data (Miro et al., 2018, p. 4), which also is the method used by Greenville water. Although categories of water use can be used (e.g., single-unit housing, multi-unit housing; commercial, industrial and agricultural), Greenville Water does not use these categories in its long-term forecasts. As described by Rinaudo (2015, p. 241) water demand forecasts are done for the short-term (hourly peak, daily peak, tomorrow, next week, next month), the intermediate term (one to ten years) and long-term (20–30 years). As indicated by the target year of 2050, our case study of Greenville Water is focused on the long term.

One of the more difficult technical barriers to constructing long-range population projections that are based on historic census data in addition to recent census data is that census geography changes from one decennial to the next and often significantly. A BG in 2010 could remain the same, it could be split into multiple smaller BGs (such as if it had experienced a lot of growth) or it could be dissolved along with other BGs around it if it is part of a greater area that is in decline. In order to account for this, the Census Bureau publishes an enormous file called a "block relationship file"—which documents every one of these geographic changes from one decade to the next (https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html).

In the case of Greenville, we chose to use 2000, 2010 and 2020 census data in the development of population projections for the target year, 2050. Thus, the first step was to re-form the 2000 and 2010 block groups and their total population data into the 2020 BG geography, which is shown in Fig. 2. For both 2000 (NGHIS, no date) and 2010 (U.S. Census Bureau, 2011), the corresponding decennial PL 94–171 redistricting file respectively, was used as the total BG population data input. Similarly, for 2020, the total BG population data from the PL 94–171 redistricting file was also employed (U.S. Census Bureau, 2022a). So, the initial projections were based on 2000 and 2010 census data in consistent 2020 census geography and launched from 2010. For reference purposes, the 2010–2020 change in population by BG is shown in Fig. 3, which also shows the boundaries of the Greenville Water Service Area.

Once the 2000 and 2010 BG geographies and data were consistent with the 2020 geography and data, we sought to create a baseline series of preliminary projections, as well as some diagnostics to assess the reliability of the 2020 Census data, which was a concern because of the introduction by the US Census

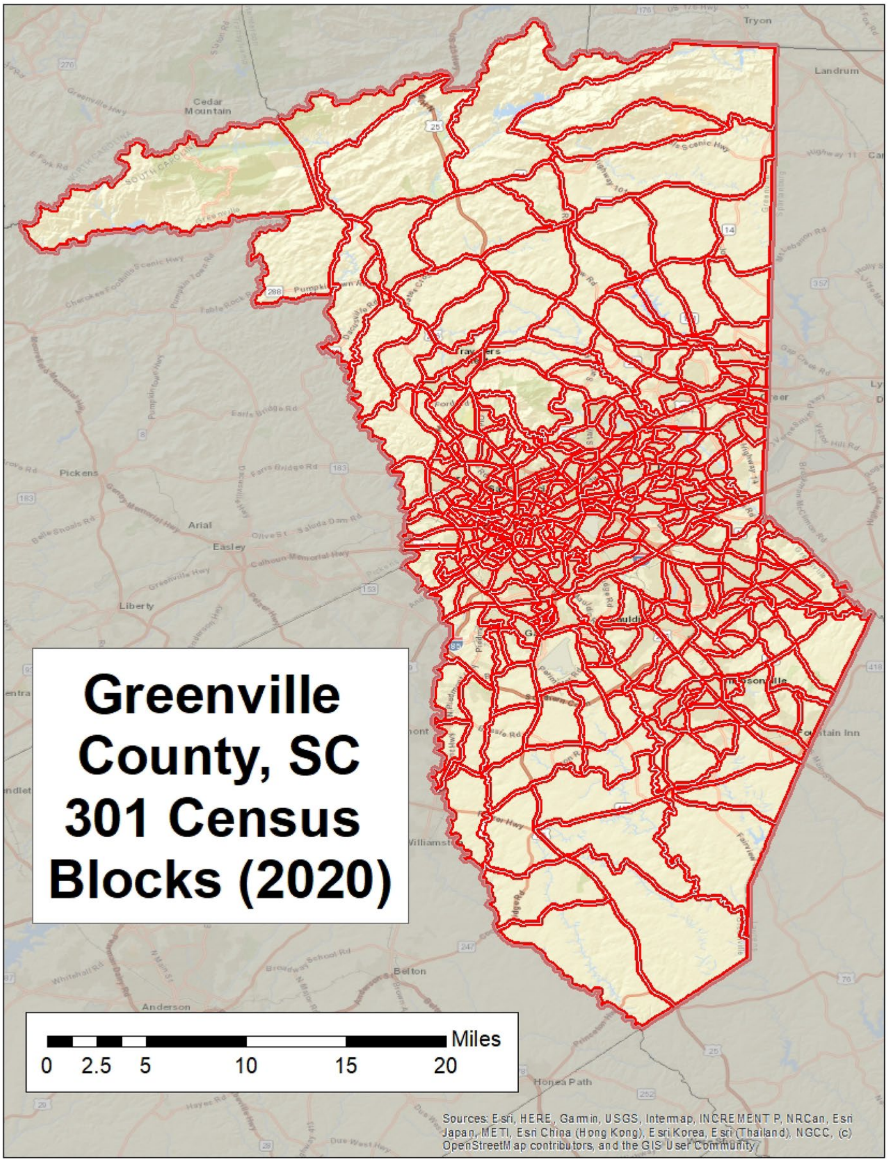

Fig. 2 Greenville County block group map

Bureau of the new Disclosure Avoidance System known as "Differential Privacy," which, as noted earlier, has been shown to induce errors in small area 2020 census data (Hauer & Santos-Lozada, 2021; Sullivan, 2020, pp. 68–81; Swanson & Bryan, 2022; Swanson & Cossman, 2021; Swanson et al., 2022; Winkler et al., 2021).

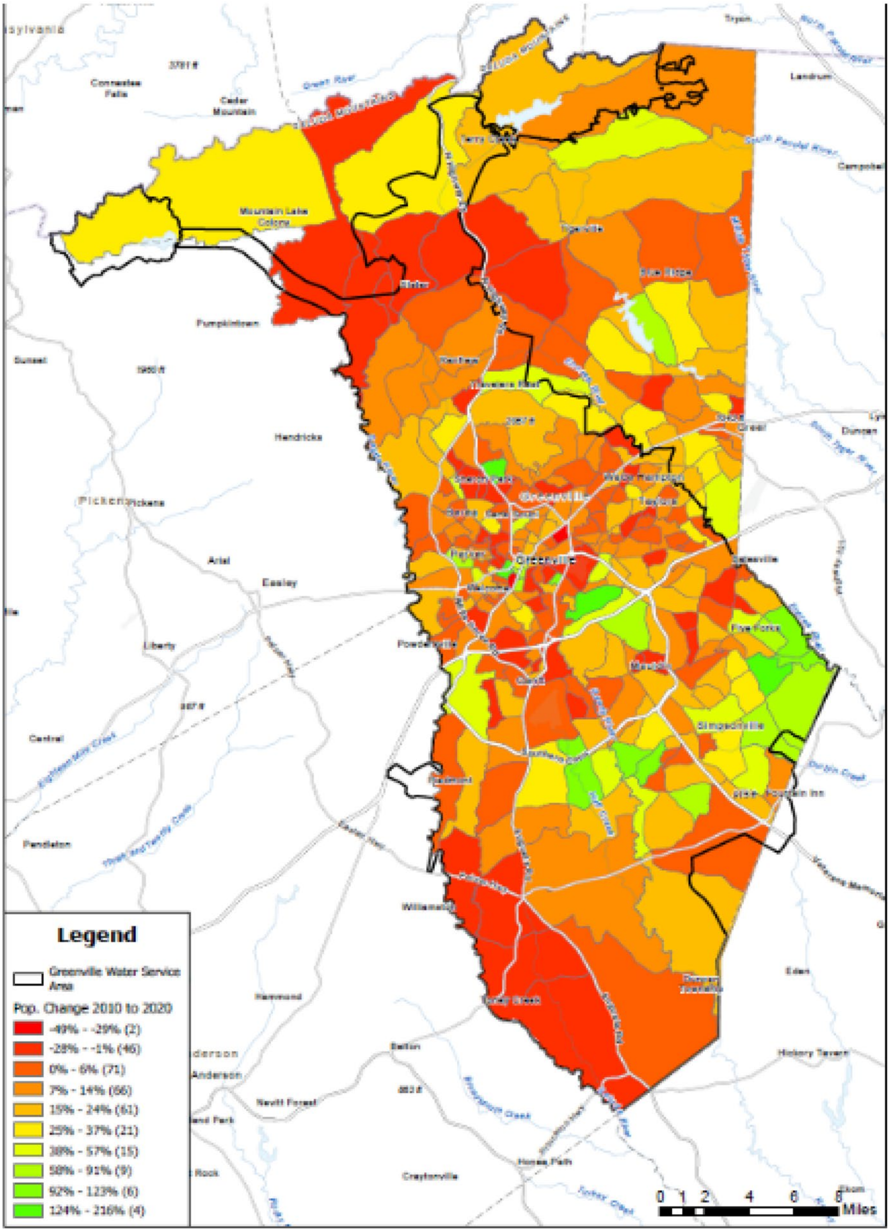

Fig. 3 Greenville County 2010–2020 population change

Three extrapolation methods were applied to the BG population totals:

1. Linear, exponential, and logistic models (Smith, Tayman, and Swanson 2013) were used to generate preliminary projections as follows

    (a) One set of linear and exponential projections were launched from 2010 using 2000-2010 linear and exponential rates of change found from 2000 and 2020 census data.
    (b) second set of linear and exponential launched from 2020 using 2010-2020 linear and exponential rates of change found from 2010 and 2020 census data.
    (c) A logistic (or constrained "S curve") model was fitted to 2000, 2010, and 2020 census data.

2. The linear and exponential models were evaluated and were observed to have generated unrealistic growth and decline in several BGs, which is why we employed several methods specific to the environment of different types of BGs. As an example of unrealistic growth, BG 450450018082 had a population of 470 in 2000 and a population of 1,580 in 2010. Using the 2000–2010 rate of geometric change and launching from 2010, its 2050 population was projected to be 102,254. As an example of unrealistic decline, BG 450450007002 had a 2000 population of 715 and a 2010 population of 518. Using the 2000–2010 rate of linear change and launching from the 2010 population, its 2050 population was projected to be -270. Given these examples, you can see why we set ceilings of 5000 and floors of zero (no negative population numbers). Following guidelines found in Swanson et al., (2010 we set these boundaries because no BG in 2010 or 2020 had a population of 5,000 or more and, obviously, no BG had a population less than zero. The logistic model results are not subject to these limits because, unlike the exponential and linear models, there are ceilings and floors are already built into them.
3. The linear and exponential projections with the ceilings and floors applied in step 2, were then used to generate preliminary BG as follows:

    (A) The arithmetic average of the logistic projections for 2030, 2040, and 2050, which, in turn, respectively, are averaged with both:
    (B) The average of the 2010 launched linear and exponential results for these same three years, 2030, 2040, and 2050 (with the ceilings of 5000 and floors of zero); and
    (C) The average of the 2020 launched linear and exponential results for these same three years, 2030, 2040, and 2050, with ceilings of 5000 and floors of zero.

    These three sets of averages yield one population projection per BG.
4. As a means of evaluating both the 2020 Census BG results (which recall may have large errors due to the application of differential privacy) and the prelimi-

nary projections, we then interpolated between the 2030 projection for each BG using the results of Step 3.B and its 2010 decennial count to obtain an estimated 2020 population for each BG, which we then compared to the official 2020 BG population. Out of 301 BGs in Greenville County, 10% had a difference of ±10% or more. Along with the finding that only 39 BGs had one or more census blocks within them where children resided with no adults present (an indicator of the effect of Differential Privacy in the absence of group quarters, Swanson & Bryan, 2022; Swanson & Cossman, 2021), this suggested that Differential Privacy was not very likely to have a deleterious effect on the BG forecasts for the district as a whole and its subareas.

5. Overall, the method we used to obtain the preliminary projections can be described as "bottom-up" (Smith et al., 2002, pp. 258–266).

Once the preliminary projections were completed, we applied expert judgment to each of the 301 BGs, keeping in mind the context of nearby BGs. The use of expert judgment is not new (Buckley & Sniezek., 1992; Dalkey, 1968; Pittenger, 1978; Roe et al., 1992; Snieczek, 1992: Swanson et al., 1995; Swanson, et al., 1997; Swanson, et al., 1998) and our application of it went as follows. The Greenville Water Engineer leading this project along with staff were familiar with the area served as the "local experts" with contracting engineers serving as a critical audience and technical support staff. The demographers presented the preliminary projections in a virtual environment along with a GIS-based set of BG maps with overlays that included recent satellite and aerial imagery. The GIS overlays (see Fig. 3 for an example_ showed the current status of land use along with environmental (e.g., ravines, creeks, ponds, wetlands, and lakes) and constructed features (quarries, highways and roads, railroad lines and yards, business, commercial and industrial complexes, college and university campuses), the history of population change, and the expected future change as indicated by the preliminary projections. Over several multiple-day periods, the team went over the GIS overlays and the data and heard the initial opinions of the local experts about the past and future of each BG. We used a categorization scheme similar to one developed by the San Diego Association of Governments for classifying the potential for population change by parcel (2008, p. 35). The coding system we developed placed each BG into one of the following seven "population growth" categories as of 2020: (1) already at build-out capacity, no further growth likely under the current land use and zoning: freeze at its 2020 number; (2) currently populated but with moderate growth likely under current land use and zoning, modify its preliminary trend for higher growth; (3) not in the Greenville Water service area, but apply expert judgment to its preliminary trend and modify if so indicated; (4) currently populated but with high growth likely under current land use and zoning; (5) rural with very low growth expected, apply expert judgment to its preliminary trend and modify if so indicated; (6) currently not growing or in decline, but zoning has changed and multi-unit, vertical housing development is either occurring or expected along with high growth, modify preliminary trend accordingly; and (7) poised for more growth, modify preliminary trend accordingly. As can be seen in Table 2, of the 301 BGs in Greenville County: (1) 118 received a classification of 1; (2) seven received a classification of 2; (3) seven received a classification of 3;

Table 2 Blockgroups by population growth category

Population category	Number of blockgroups
1	118
2	7
3	7
4	0
5	17
6	38
7	114

See text for the category descriptions

(4) zero received a classification of 4; (5) 17 received a classification of 5; (6) 38 received a classification of 6; and (7) 114 received a classification of 7.

These same classifications were used as the forecast went to 2030 and 2040, with modifications as needed (e.g., as of 2030, if a BG reached build-out capacity and was not likely to grow further, it was re-classified accordingly). In the few cases where there was a majority but not a unanimous decision into which category a given BG should be placed, it was resolved by placing it into the nearest category that allowed for further growth. The rationale was that Greenville Water would rather err on the side of over-building than under-building. When this process was completed, nearly half of the BGs were classified as being effectively "done" and unlikely to change size in the future. This allowed us to focus on the remaining BGs regarding how much growth or decline was expected.

Once this process was completed, the demographers revised the preliminary projections accordingly and labeled them as "provisional projections." Of the 301 BGs, 25 were selected for further review, which resulted in revisions. All of these 25 BGs are found among the 38 members of Group "6." The revised provisional BG projections were summed, respectively, for 2030, 2040, and 2050 to obtain county totals for 2030, 2040, and 2050, which were then compared to independently generated county population projections for these same years. As discussed in the following section, the BG sums were found to fall within the range of the independently generated county population totals and with no more changes to the BG projections, the results were classified as the "final forecasts." Including preparation, this entire expert review process took about 150 person-hours and occurred over several multi-day periods using a virtual meeting (zoom) application.

3 Results

Using the data and methods described in the preceding section, provisional forecasts of the BGs within Greenville County were developed. These forecasts were informed by expert judgment as described earlier, which is important to use when forecasting the population of small areas such as BGs for a specific geographic area such as the Greenville Water (Swanson et al., 1998, 2010).

As one of several validity checks, we compared these county totals not only to those found from other sources, including those shown in the 2016 Master Plan (Fig. 4.2) (2016), but also to a "Hamilton-Perry" projection produced by the demographers and described in footnote 3 of Table 1, which contains the results of these comparisons. It appears from the data in Table 1 that this update is timely, given our county total of 837,838 for 2050 compared to the numbers found in Fig. 4.2 of the 2016 Master Plan, where: (1) the 2009 Master plan has a 2035 County Population of approximately 570,000; (2) the 2014 Keowee-Taxaway relicensing project has a 2040 county population of approximately 675,000; (3) the 2014 ACOG has a 2040 county total of approximately 670,000; and (4) the "extrapolated" *State of SC* results, which has a county total of 812,519 by 2050.

An important component of these comparative projections is the "Hamilton-Perry Method" that we produced, which is based on cohort change ratios and is directly linked to the fundamental theorem of population change (Baker et al., 2017, pp. 251–252). This serves to place our forecasts on a theoretical foundation and conforms to the arguments made by Burch (2018) in regard to demographic theory, a point to which we return in the following section.

The 2050 Greenville County total for the provisional "informed" forecast is between the highest of the independently generated county 2050 totals we generated using ARIMA (which we generated using the NCSS statistical software system release 12.0.4 in the of a "log10(pop)-trend" model with one parameter, – 0.502 and (p,d,q) = (1,2,0) = 1 order of auto-regression, 2 orders of differencing, and no moving average) which is 859,734, and the State of South Carolina projection which terminates in 2035, but for which we applied the geometric rate of change from 2025 to 2035 to generate a projection to 2040 and ultimately, 2050, where have a projection of 812,699. The lowest independently generated 2050 county total we did was 759,090, which used the Hamilton-Perry Method and 2010–2015 input data from the US. Census Bureau's "Population Estimates Program" (2022b).

The increase of 152,840 people in the year 2050 that is spread over all of the 301 census block groups in going from the preliminary projection (675,626) to the final informed forecast (828,467) is largely generated by 25 (of 301) BGs, which expert judgment determined were currently poised to "take off" in terms of population growth. As noted earlier, all of these BGs were categorized as Group "6." Having this much change (23%) generated by such a small number of BGs is consistent with the findings of Baker et al. (2021), who found in an ex post facto evaluation of 2010 census tract forecasts for the U.S. as a whole that the bulk of the errors for 65,221 census tracts was driven by high growth rates in one percent of the tracts between 2000 and 2010. In the case at hand, just under 1 percent (0.83%) of the BGs is driving much of the 23% increase of the total 2050 Greenville County population found in going from the preliminary projection to the final, informed forecast. Had the local expert process not been used, the preliminary BG projections would have missed this expected growth because there was no history of such growth in the last 20 years of census data for these BGs.

As examples of the results of the local expert judgment, for selected BGs, we compare the results of the preliminary projections to the final forecast results. The seven BGs selected represent, respectively, each of the seven categories used to

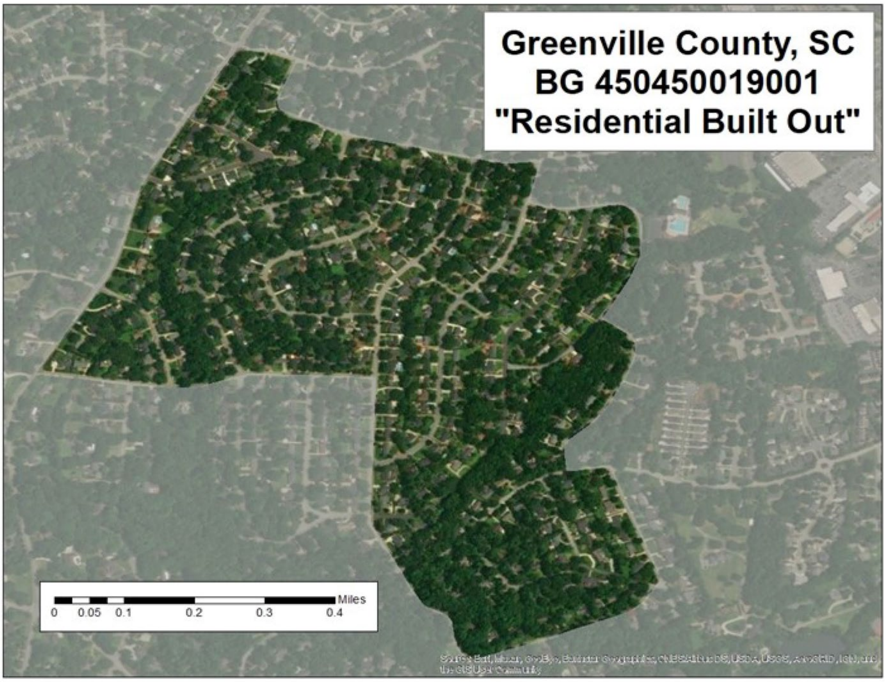

Fig. 4 GIS Overlay example of a BG placed in Category 1, "No Growth Expected." See text in the Results section

classify them during the local expert process. Shortly, we discuss two of these seven BGs for which we provide the corresponding "screen shots" (Fig. 4, "category 1, no growth expected" and Fig. 5, "category 6, high growth expected") as examples of what the local experts viewed in making the decision to classify a given BG into one of the seven categories.

Coded as a "1" (freeze the population at either the 2020 or preliminary level), Blockgroup 4,504,500,019,001 had a 2020 population of 2,463, resulting in a preliminary 2050 population of 2643 and a final forecasted 2050 population of 2643 (See Fig. 4);

Coded as a "2" (modify its preliminary trend for higher growth), Blockgroup 450,450,002,001 had a preliminary 2050 population of 1836, with a final forecasted 2050 population of 2139;

Coded as a "3" (Not in the Service Area, but modify its preliminary trend if expert judgment indicates), Blockgroup 450,450,040,031 had a preliminary 2050 population of 2215 with a final forecasted 2050 population of 2140;

Coded as a "4" (currently populated but with high growth likely under current land use and zoning), No Blockgroups were found to fit this category;

Coded as a "5" (rural with very low growth historically, but apply expert judgment to its preliminary trend and modify accordingly if required), Blockgroup 450,450,039,021 had a preliminary 2050 population of 2202 with a final forecasted 2050 population of 2700;

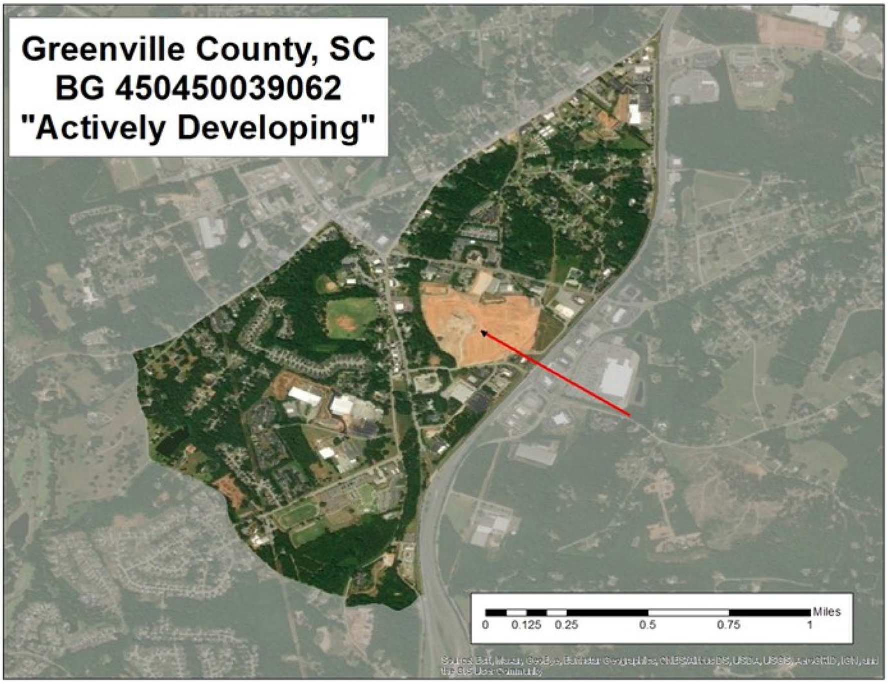

Fig. 5 GIS Overlay Example of a BG placed in Category 6, "High Growth Expected." See text in the Results section

Coded as a "6" (currently not growing or in decline, but zoning has changed and vertical housing development is either occurring or expected, along with high growth, modify preliminary trend accordingly), Blockgroup 450,450,039,062 had a preliminary 2050 population of 1633, with a final forecasted 2050 population of 5619 (See Fig. 5); and.

Coded as a "7" (poised for more growth, whether modest or substantial, and modify preliminary trend accordingly), Blockgroup 450,450,038,023 had a preliminary 2050 population of 1503, with a final forecasted population of 1708.

The two examples shown in Figs. 4 and 5, respectively, represents the low and high ends of expected BG population growth over the forecast horizon to 2050. As can be seen in Fig. 4 ("category 1, no growth expected"), the homes are well-established and the local experts knew that there had been no zoning changes that would transform this from a low density, single-household residential neighborhood to a high density, multi-household residential neighborhood through the end of the forecast horizon. In Fig. 5, ("category 6, high growth expected"), clearly visible is the cleared and platted area for the high density, multi-household units that the local experts knew were coming due to zoning changes and other decisions that were recently made., which, once in place, would likely remain until the end of the forecast horizon.

4 Discussion

There will be change between 2020 and 2050 in the 118 BGs where the future populations were frozen at the 2020 level. However, the expert judgment is that the change will be minimal and not worth the effort to model. This decision allowed the time to focus attention on the BGs where moderate and, especially, substantial change was believed likely. While this reduced the person-hours of labor, the hybrid forecasting process was still labor-intensive, which experience suggests that it is worthwhile if a reasonable level of accuracy is desired for small area forecasts (Swanson et al., 2010). Swanson et al. (2010) conducted an ex post facto evaluation of the accuracy of small area forecasts informed by local experts in the Hillsborough School District project and found that the correct decisions were made concerning the need for new construction within the forecast horizon, namely that while a new high school and elementary school were needed, a new middle school was not. This is an example of the level of accuracy required for this study. That is, it does not need to be accurate to the point of perfection (Swanson et al., 1996); it only needs to be sufficiently accurate that the right decisions are made regarding the construction and placement of new infrastructure, which in this case is for a water district. Finding that there was a difference of 23% (152,840) for the 2050 county total population in going from the preliminary projections (675,626) to the final, informed forecasts (828,467), suggests that if the preliminary set of BG projections were used for its planning needs, Greenville Water is likely to have under-estimated the demand for water in its service area by about 18 percent as of the year 2050. Nearly as important, we identified block groups that had grown significantly in the past but have since stabilized or even begun to decline somewhat. Had these BGs been allowed to continue apace, the size and drivers of change in the county would have changed dramatically.

Arguably, the strongest feature of the process described here is that a combination of different quantitative approaches used in conjunction with knowledgeable judgment can be expected to provide forecasts that are sufficiently accurate to inform decision-making without being prohibitively expensive. Another feature that serves to strengthen the process is the use of the Cohort Component Method in four of the six county population projections and the Cohort Change Ratio Method as the basis of the fifth of these six projections (See Table 1), which means that five of the six methods we employ for this purpose are, as we discussed earlier, directly linked to the fundamental elements of demographic theory. In turn, the linkage from these county level projections to the BG projections provides the latter with an indirect, but important, connection to this same foundation. Continuing along this line, in regard to one of the two simple extrapolative projection methods we employ in this case study, the exponential model, Burch (2018 p. 45) notes that it also is based on demographic theory, namely the theory of how a population grows. This places our BG population projections even closer to the key theoretical elements we discussed in the "Strengths and Weaknesses" section that serve to make a forecast stronger in that they correspond to the dynamics by which a population moves forward in time.

There are also additional benefits to the hybrid forecasting process described here. These include the development of a set of potential "opinion leaders" in the form of the local experts who serve as informal diffusion channels regarding recommendations and their rationales before the diffusion through more formal channels such as public hearings. This represents a valuable communication channel. In addition, the local experts represent a community resource in the form of a group very knowledgeable not only about the likely water demands and demographic future of Greenville County, SC but also about a range of forecasting methods, and their respective strengths and weaknesses. Similarly, the demographers become knowledgeable about local area conditions and sources of information that serve to make them more proficient in the craft of small area forecasting. We believe that other water districts can benefit from the conceptual system that underlies the specific process used for the "Greenville Water" population forecasts. We recommend that it be considered in water districts facing the need to assess capital facilities and develop or revise supply and distribution infrastructure.

In this case study, all of the team members and the client worked with a common understanding that the forecast population will be different than the actual 2050 population. In light of both the "errors" in the projection methods and the "errors in the expert judgment" applied to them, we did, however, ask the client directly which type of bias would be preferred in our population forecast, upward or downward? The client responded that an upward bias would be preferred because a downward bias would likely necessitate the need for another long-term plan sooner than anticipated whereas an upward plan would put in place infrastructure that might be premature but would eventually be used. As discussed in the preceding section, the experts, followed this guidance when there was there was a difference of opinion on the future population of a given BG.

Barring a "Black Swan" event (Taleb, 2010), the demographers who participated in this case study expect that the error in the final 2050 population forecast for the district as a whole and most of the 301 BGs will be in the range of $\pm 5\%$, with a handful of BGs much higher ($> \pm 20\%$), That is, with a small set of extreme errors relatively similar that found for census tracts by Baker et al. (2023) when they used a machine learning approach to reduce the number of extreme projection errors. In the absence of the expert judgment process, the demographers believe the error in the final 2050 population forecast for the district as a whole and most of the 301 BGs would have been in the range of $\pm 15\%$, with a handful of BGs very much higher ($> \pm 50\%$). That is, with a small set of extreme errors relatively similar to that found for census tracts by Baker et al. (2021) in the absence of using a machine learning approach to reduce the number of extreme projection errors.

In closing, we observe that the release of the new 2020 census data enabled the decennial exercise of political apportionment and redistricting. With that enormous body of work largely done, applied demographers are now turning their attention to an equally large body of post-censal applied demography projects such as the case study presented here. An important and significant share of those projects is the exercise of creating small area population projections, which are used for public policy and infrastructure planning. Census data have never been well suited for small area population projections "out of the box"—but this is particularly true for the 2020

Census, which in addition to the usual decennial problems included political interference by the Trump administration, implementation of a new (online) response system, dealing with a pandemic, curtailment of the non-response follow-up period, and the Bureau's choice of differential privacy as its new disclosure avoidance system (Hauer & Santos-Lozada, 2021; Hotz & Salvo, 2022; O'Hare, 2020; Ruggles & Van Riper, 2022; Swanson, 2021; Swanson & Cossman, 2021; Swanson et al., 2022; Winkler et al., 2021). Hopefully, this paper also provides ideas about how to work around the additional problems found in the 2020 census.

Acknowledgements We are grateful to Jeff Tayman for comments on an earlier draft and Ron Provost at Georgetown University for providing information on Blockgroups in Greenville County that are likely affected by Differential Privacy. We also thank the reviewers for their comments on an earlier draft.

Author contributions All data are secondary and all authors contributed to this paper.

Funding Greenville Water funded this research.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Baker, J., Swanson, D. A., & Tayman, J. (2021). The accuracy of hamilton-perry population projections for census tracts in the United States. *Population Research and Policy Review, 40*(6), 1341–1354.

Baker, J., Swanson, D. A., & Tayman, J. (2023). Boosted regression trees for small area population forecasting. *Population Research and Policy Review, 42*, 51. https://doi.org/10.1007/s11113-023-09795-x

Baker, J., Swanson, D. A., Tayman, J., & Tedrow, L. (2017). *Cohort change ratios and their applications*. Springer Press.

Buckley, T., & Sniezek, J. (1992). Passion, preference, and predictability in judgment forecasting. *Psychological Reports, 70*, 1022.

Burch, T. (2018). *Model-based demography: Essays on integrating data, technique, and theory*. Demographic Research Monographs. Springer.

Butler, D., & Ali Memon, F. (2005). *Water demand management*. IWA Publishing.

Dalkey, N. (1968). *Predicting the future*. Rand Paper (https://www.rand.org/pubs/papers/P3948.html)

Green, K., & Armstrong, J. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research, 68*, 1678–1685.

Greenville Water (2016). *Water resources master plan*. Greenville Water. Greenville, SC.

Hauer, M., & Santos-Lozada, A. (2021). Differential privacy in the 2020 census will distort COVID-19 rates. *Socius*. https://doi.org/10.1177/2378023121994014

Hogarth, R., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science, 27*(2), 115–138.

Hotz, V., & Salvo., J. (2022). A chronicle of the application of differential privacy to the 2020 Census. *Harvard Data Science Review* (Special Issue 2). https://hdsr.mitpress.mit.edu/pub/ql9z7ehf/release/7?readingCollection=63678f6d.

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin books.

Miro, M., Groves, D., Catt, D. & Miller, B. (2018). *Estimating future water demand for San Bernardino Valley Municipal Water District*. RAND WR-1288-SBVMWD. Santa Monica, CA.

NGHIS. (no date). *NHGIS dataset groups, 2000_PL9417 redistricting data summary file,* https://www.nhgis.org/overview-nhgis-datasets#dsg44.

O'Hare, W. (2020). The politicization of the 2020 Census. *PAA Affairs,* https://higherlogicdownload.s3.amazonaws.com/POPULATIONASSOCIATION/3e04a602-09fe-49d8-93e4-1dd0069a7f14/UploadedImages/Documents/PAA_Affairs/PAA-Fall_20_.pdf .

Petropoulos, F., et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting, 38*, 705–871.

Pittenger, D. (1978). The role of judgment, assumptions, techniques, and confidence intervals in forecasting population. *Socio-Economic Planning Sciences, 12*, 271–276.

Rinaudo, J., et al. (2015). Long-term water demand forecasting. In Q. Grafton (Ed.), *Understand and managing urban water in transition* (pp. 239–268). Springer Press.

Roe, L., Swanson, D., & Carlson, J. (1992). A variation of the housing unit method for estimating the population of small, rural areas: A case study of the local expert procedure. *Survey Methodology, 18*(1), 155–163.

Ruggles, S., & Van Riper, D. (2022). The role of chance in the census bureau database reconstruction experiment. *Population Research and Policy Review, 41*, 781–788. https://doi.org/10.1007/s11113-021-09674-3

San Diego Association of Governments. (2008). *2030 Regional growth forecast update: Process and model documentation*. San Diego Association of Governments. San Diego, CA.

Smith, S., Tayman, J., & Swanson, D. (2002). *State and local population projections: Methodology and analysis*. Kluwer Academic Publishers.

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

Sniezek, J. (1992). Groups under uncertainty: An examination of confidence in group decision-making. *Organizational Behavior and Human Decision Processes, 52*(1), 124–155.

Sullivan, T. (2020). *Census 2020: Understanding the issues*. Springer Press.

Swanson, D., Roe, L., Carlson, J. & Williams, C. (1995). Estimating the population of rural communities by age and gender: A case study of the local expert procedure. *Small Town* (May-June):14–21.

Swanson, D. (2021). Broadband access during a pandemic: 2020 Census results for the Hopi and Lummi Reservations. *PAA Affairs.* 11–29–2021, https://www.populationassociation.org/blogs/david-swanson/2021/11/29/broadband-access-during-a-pandemic-2020-census?CommunityKey=a7bf5d77-d09b-4907-9e17-468af4bdf4a6.

Swanson, D., & Cossman, R. (2021). The effect of the differential privacy disclosure avoidance system proposed by the Census Bureau on 2020 Census products: Four case studies of Census blocks in Mississippi. Presented at the Joint Statistical Meeting, American Statistical Association, Seattle, WA. https://www.amstat.org/meetings/jsm/2021/onlineprogram/AbstractDetails.cfm?abstractid=317734.

Swanson, D., & Bryan, T. (2022). Repurposing record matching algorithms to identify blocks and block groups affected by differential privacy: Progress report on a pilot project. Presented at the 2022 small area estimation conference, session on challenging problems from SAE and modern data science, May 26. https://sae2022.org/public/assets/files/DavidSwanson.pdf.

Swanson, D., Bryan, T. & Sewell, R. (2022). The effect of the differential privacy disclosure avoidance system proposed by the Census Bureau on 2020 Census products: Four case studies of census blocks in Alaska. In *Proceedings of the symposium on data science and statistics*, American Statistical Association, Alexandria, VA. pp. 2058–2062. Available at https://www.populationassociation.org/blogs/paa-web1/2021/03/30/the-effect-of-the-differential-privacy-disclosure.

Swanson, D. (2022). Forecasting a tribal population using the cohort-component method: A case study of the Hopi. *Population Research and Policy Review, 41*, 1831–1852.

Swanson, D., Burch, T., & Tedrow, L. (1996). What is applied demography? *Population Research and Policy Review, 15*(5–6), 403–418.

Swanson, D., Hough, G., Lycan, D., & Sharkova, I. (1997). Forecasting for school attendance zone changes: Merging GIS and demographic methods with local expert judgment. *Market and Demographic Analysis, 3*(6), 23–31. (in Chinese).

Swanson, D., Hough, G., Lycan, R., Clemans, C., & Rodriguez, J. (1998). Merging methods and judgment for K-12 enrollment forecasting. *Educational Research Service Spectrum, 16*(Fall), 24–31.

Swanson, D., & Morrison, P. (2010). Teaching business demography using case studies. *Population Research and Policy Review., 29*(1), 93–104.

Swanson, D., Schlottmann, A., & Schmidt, R. (2010). Forecasting the population of census tracts by age and sex: An example of the hamilton-perry method in action. *Population Research and Policy Review, 29*(1), 47–63.

Swanson, D., & Tayman, J. (1995). Between a rock and a hard place: The evaluation of demographic forecasts. *Population Research and Policy Review, 14*, 233–249.
Swanson, D., & Tayman, J. (1996). On the utility of population forecasts. *Demography, 33*(4), 523–528.
Swanson, D., Tedrow, L., & Baker, J. (2016). Exploring stable population concepts from the perspective of cohort change ratios: Estimating the time to stability and intrinsic r from initial information and components of change. In R. Schoen (Ed.), *Dynamic demographic analysis* (pp. 227–258). Springer Press.
Taleb, N. (2010). *The black swan: The impact of the highly improbable* (2nd ed.). Penguin.
Texas Water Development Board. (1997*). Water for Texas: A consensus-based update to the state water plan: Vol. II, technical planning appendix. Doc # GF-6–2.* Austin, TX. http://www.twdb.state.tx.us/.
Texas Water Development Board. (2021). *2021 Regional water plan: Population and water demand projections,* https://www.twdb.texas.gov/waterplanning/data/projections/2022/popproj.asp.
Theocharis, Z., & Harvey, N. (2019). When does more mean worse? Accuracy of judgmental forecasting is nonlinearly related to length of data series. *Omega, 87*, 10–19.
U.S. Census Bureau. (2011). *2010 Census: Redistricting file (Public Law 94–171) Dataset,* https://www.census.gov/data/datasets/2010/dec/redistricting-file-pl-94-171.html.
U.S. Census Bureau. (2022a). 2020 Census redistricting data (P.L. 94–171) Summary files, https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files/2020.html#P1.
U.S. Census Bureau. (2022b). *Population and housing unit estimates datasets,* https://www.census.gov/programs-surveys/popest/data/data-sets.html.
Winkler, R. J., Butler, K. C., & Egan-Robertson, D. (2021). Differential privacy and the accuracy of county-level net migration estimates. *Population Research and Policy Review*. https://doi.org/10.1007/s11113-021-09664-5
Zellner, M., Abbas, A., Budescu, D., & Galstyan, A. (2021). A survey of human judgement and quantitative forecasting methods. *Royal Society Open Science, 8*, 201187. https://doi.org/10.1098/rsos.201187

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

David Swanson[1,2,3] · Tom Bryan[4] · Mark Hattendorf[5] · Kelly Comstock[6] · Lauren Starosta[6] · Robert Schmidt[5]

✉ David Swanson
dswanson@ucr.edu

1 University of California Riverside, Riverside, CA 92521, USA

2 Population Research Center, Portland State University, Portland, OR 97201, USA

3 Center for Studies in Demography and Ecology, University of Washington, Seattle, WA 98195, USA

4 Bryan Geodemographics, Midlothian, VA 23114, USA

5 Greenville Water, Greenville, SC 29601, USA

6 Brown and Caldwell, Columbia, SC 29210, USA