PROBABILISTIC INTERVALS AROUND POPULATION FORECASTS: A NEW APPROACH WITH A SUBNATIONAL EXAMPLE USING WASHINGTON STATE COUNTIES

David A. Swanson[1] – **Jeff Tayman**[2]

Abstract

Population forecasts produced by governments at all levels are used in the public sector, the private sector, and by researchers. They have been primarily produced using deterministic methods. This paper shows how a method for producing measures of uncertainty can be applied to existing subnational population forecasts while meeting several important criteria, including the concept of utility. The paper includes an assessment of the efficacy of the method by: (1) examining the change in uncertainty intervals it produces by population size and population growth rate; and (2) comparing the width and temporal change of the uncertainty intervals it produces to the width and temporal change of uncertainty intervals produced by a Bayesian approach. The approach follows the logic of the Espenshade-Tayman method for producing confidence intervals in conjunction with ARIMA equations to construct a probabilistic interval around the total populations forecasted from the Cohort Component Method, the typical approach used by demographers. The paper finds that population size and population growth rate are related to the width of the forecast intervals, with size being the stronger predictor, and the intervals from the proposed method are not dissimilar to those produced by a Bayesian approach. This approach appears to be well-suited for generating probabilistic population forecasts in the United States and elsewhere where these forecasts are routinely produced. It has a higher level of utility, is simpler, and is more accessible to those tasked with producing measures of uncertainty around population forecasts.

Keywords: ARIMA, Bayesian Methods, Cohort Forecasting Methods, Espenshade-Tayman Method, Forecast Uncertainty, Utility
Demografie, 2025, 67(3): 124–147
DOI: https://doi.org/10.54694/dem.0365

1) Population Research Center, Portland State University, Center for Studies in Demography and Ecology, University of Washington, Department of Sociology, University of California Riverside. Contact: dswanson@ucr.edu.
2) Tayman Demographics, 2142 Diamond St., San Diego, CA 92109. Contact: jtayman@san.rr.com.

INTRODUCTION

As *Wu et al.* (2023) observe, population forecasts produced by governments at all levels are used in the public sector, the private sector, and by researchers. However, until recently, these widely used forecasts have primarily been produced using deterministic methods in conjunction with the Cohort Component Method (CCM) and its algebraic equivalent, the Cohort Change Ratio (CCR) approach (see Appendix A) – a practice consistent with an observation made by *Baker, Alcantara, and Ruan* (2011: 10): "Demographic modeling occurs without consideration of statistical uncertainty." They noted that this oversight applied specifically to population forecasting. Regarding subnational forecasting, we find that the exception to their observation consists of five studies that have presented methods for developing probabilistic population forecasts: *Cameron and Poot* (2011); *Swanson and Beck* (1994); *Swanson and Tayman* (2014); *Wilson* (2012); and *Yu et al.* (2023). Notably, four of these studies link probabilistic uncertainty to the CCM approach (*Cameron – Poot,* 2011; Wilson, 2012; Yu et al., 2023) or the CCR approach (*Swanson – Tayman,* 2014). The linkage found in these four studies is significant because it means that the measures of uncertainty are linked to the fundamental demography equation, whereby a 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: 251–252).

The fundamental equation is the cornerstone of demographic theory (*Canudas-Romo et al.,* 2008; Swanson et al., 2023) and the foundation upon which the CCM rests (*Baker et al.,* 2017: 23–24; Burch, 2018). A probabilistic approach to population forecasting based on this theoretical foundation yields benefits not found in methods lacking this foundation (e.g., *Burch,* 2018; *Land,* 1986). This observation is also consistent with one made by *Swanson et al.* (2023), who argue that a given method's strengths and weaknesses largely stem from four sources: (1) its correspondence to the process 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.

Like their counterparts in the private sector and at the national level, state and local demographers are constrained by resources and time. Because of these constraints, *Tayman and Swanson* (1996) pointed out the importance of considering the concept of utility in producing population forecasts. *Swanson, Burch, and Tedrow* (1996) added more specificity to this issue by introducing the "triple constraint" perspective, which can be applied to population forecasting:

1. Performance specification – the explanatory/predictive precision sufficient to support a given decision-making situation.
2. Time – the schedule requirements under which the performance specification must be accomplished.
3. Resources – the budget requirements under which the performance specification must be accomplished.

The performance specification is directly related to the four sources identified by *Swanson et al.* (2023), as well as the strengths and weaknesses of a given approach to forecasting, such as the time and resource specifications.

It is important to note that population forecasting is considered to be part of applied demography (*Swanson – Burch – Tedrow,* 1996), where problems come not from demographic theory or empirical research traditions but from a person (or set of persons) in government, business, or some other organizational sector who needs demographic analysis to assist him or her in making good, informed decisions. A corollary identified by *Swanson, Burch, and Tedrow* (1996) is that the decision-making process is client-driven in that the definition of the problem and an adequate answer are determined primarily by the decision-maker, not by the demographer or demographic research traditions. This corollary means that the primary audience for population forecasts comprises decision-makers and their constituents, not professors and academic researchers – two groups to which the triple constraint perspective applies, but differently. Professors and academic researchers pursue ever-improved knowledge, more precise and reliable measurements, better theoretical systems, and more refined techniques. They see costs and time as constraints to overcome to achieve high performance specification levels. Decision-makers are interested in acquiring the minimal amount of information needed to make correct decisions. They want to optimize the performance specification within the constraints generated by time and resource limitations. This latter view underlies the approach to generating uncertainty information we present here.

Turning to the five exceptions, we start with *Cameron and Poot* (2011), who developed a stochastic method for subnational population forecasts. They applied it to five demographically distinct administrative areas within the Waikato region of New Zealand. The uncertainty measures found in this approach are generated around the components of population change, which means this approach is linked to the fundamental population equation. Their results are compared to official subnational deterministic forecasts, which revealed the instability of migration as a component of population change.

Swanson and Beck (1994) proposed a lagged regression-based method to generate short-term county population forecasts. It is based on modifying the ratio-correlation method of population estimation and partly on earlier work by *Swanson* (1989). The modified ratio-correlation method produces forecasts without requiring substantial data and intensive intellectual labor inputs. Tests found that this approach delivered accurate forecasts (*Swanson – Beck,* 1994).

Swanson and Tayman (2014) examined state-level forecasts using a lagged regression approach in conjunction with the Cohort Change Ratio Method (CCR). As *Baker et al.* (2017) discuss, the CCR approach is algebraically equivalent to the CCM approach but uses cohort change ratios to capture mortality and migration; and to capture fertility, it uses child-adult ratios. This equivalency means the CCR approach is linked to the fundamental population equation. *Swanson and Tayman* (2014) found that the uncertainty measures associated with their lagged-regression CCR approach were not too wide in that they captured reported totals and age groups in accordance with expectations.

Wilson (2012) used the empirical approach to develop uncertainty measures for the subnational forecasts he constructed in Australia. It is based on empirical analyses of errors from past forecasts (*Smith – Tayman – Swanson,* 2013; Stoto, 1983). Although not a formal method for generating uncertainty measures, it is valid and Wilson applied it to CCM forecasts, which means that they are linked to the fundamental population equation. Importantly, Wilson's application provides uncertainty measures for age groups. This work has a precedent in which *Wilson* (2005) applied time series methods and judgment to develop uncertainty measures for New Zealand's national population forecasts.

Using the 39 counties of Washington state as an example, *Yu et al.* (2023) show that a Bayesian approach can be used in conjunction with the CCM to provide probabilistic county-level population forecasts. Like *Wilson* (2012), this approach measures uncertainty for age groups. To our knowledge, this is the first application of Bayesian inference to the CCM approach for projecting county populations. It is a seminal contribution. However, using experience as a guide, we also believe that it will take time for this approach to be widely adopted by the state and local demography communities, in part because Bayesian inference can be complex, effortful, opaque, and even counter-intuitive (*Goodwin,* 2015).

This paper adds to the sparse literature on subnational probabilistic population forecasting that has been reviewed here by describing a new approach for constructing uncertainty measures that is relatively simple and can be linked directly to either the CCM or CCR approach. Importantly, unlike Bayesian inference, we believe this new approach is likely to meet essential evaluation criteria routinely used by state and local demographers (*Smith – Tayman – Swanson,* 2013: 301–322), such as low production costs (particularly staff time), application and explanation ease, a high level of face validity, and intuitive. The approach we propose employs the ARIMA method in conjunction with work by *Espenshade and Tayman* (1982) to translate the uncertainty information in the ARIMA method's forecast to the population forecast provided by the CCM approach.

DATA

Following *Yu et al.* (2023), we use data for the 39 Washington state counties to demonstrate our new approach to placing probabilistic intervals around forecasts produced using CCM. We also evaluate our new approach by comparing its results to those reported by *Yu et al.* (2023). Figure 1 provides a map of Washington's counties. According to the 2020 decennial census, Washington state had a population of 7.7 million, the thirteenth largest state in the U.S. Its growth rate between 2010 and 2020 was 14.6%, the seventh fastest growth rate among states. Washington's 39 counties reflect a broad range of population sizes and growth rates and provide a diverse data set for evaluating the new approach, which places intervals around CCM forecasts. According to the 2020 decennial census, the average county population was 197,571, ranging from 2,286 in Garfield County to 2,267,675 in King County. Forty-six percent of the counties were smaller than 50,000, and only three counties exceeded 500,000 persons. The average county growth rate from 2010–2020 was 9.5%, ranging from –4.9% in Ferry County to 23.8% in Franklin County. Eight counties (20%) experienced growth rates below 5%, and seven counties (18%) experienced growth rates above 15%.

Figure 1 **Washingon State Counties**

Source: Washington Association of County Officials (https://countyofficials.org/192/County-Map)

We use annual intercensal estimates from 1960 to 2020 produced by the Forecasting Division of the Office of Financial Management (OFM) to implement the ARIMA model (*Washington,* 2024). Intercensal estimates are developed between census years and are considered more accurate than other estimates because either decennial census counts by the U.S. Census Bureau or state-certified special census counts on both sides bracket them. The intercensal estimates are based on the housing unit method (e.g., *Swanson – Tayman,* 2012: 137–164). The housing unit method assumes that the change in the number of people varies with the change in the number of housing units and counts of the population living in group quarters facilities, as reported to OFM by local governments and institutions over the decade.

We launched the forecasts from 2020 to match the launch year (2020) of the 2022 Growth Management Act (GMA) county forecasts (*Washington,* 2022). These GMA forecasts contain scenario-based, not probability-based, intervals around the medium forecast and have a 30-year forecast horizon to 2050.

Population forecasts for the GMA are produced every five years and developed with county officials. Directed by state statute (House Bill 1241), OFM prepares a reasonable range of possible population growth for Washington state's counties. County officials, also by law, are responsible for selecting a 20-year GMA planning target from within the range of high and low prepared by OFM.

Our approach follows the logic of the Espenshade-Tayman method (*Espenshade – Tayman,* 1982) for producing confidence intervals around postcensal population estimates by age. Their method employs time-series regression equations to construct probabilistic intervals around age-specific death rates over a postcensal estimation period. These results, combined with the number of deaths during this period and the most recent census counts, were translated into confidence intervals around the corresponding estimated age structure. Our use of the Espenshade-Tayman method is not unique. It has been employed by *Swanson* (1989) and *Roe, Swanson, and Carlson* (1992) in demographic applications.

Our approach uses ARIMA models to generate confidence intervals around population densities.[3] We use "density" because the *Espenshade-Tayman* (1982) method for translating uncertainty information does so from an estimated "rate," which in this case is the "rate" of population density. Other "rates" could be used, such as the ratio of the population to the number of housing units. However, using the land area as the denominator provides a virtually constant denominator over time, thereby reducing the effort in assembling the "rate" data. It also serves as a stabilizing element regarding the use of ARIMA in that it dampens the effect of short-term population fluctuations more effectively than, say, housing units, which also can fluctuate over time and not always in concert with population fluctuations.

Three steps are needed to generate a confidence interval around the GMA point forecast produced by the CCM, which we label here as "GMAPOP." First, ARIMA models produce a point forecast for population density (which we label here as "PFPD"), along with a lower limit (which we label here as "LLPD"), and an upper limit (which we label here as "ULPD") for each county and Washington state. Second, relative differences (proportions) are determined for each lower limit (which we label here as "RLLPD") and upper limit (which we label here as "RULPD"). These relative differences are found as follows:

$$\text{RLLPD} = (\text{LLPD} - \text{PFPD}) / \text{PFPD} \text{ and} \quad (1)$$

$$\text{RULPD} = (\text{ULPD} - \text{PFPD}) / \text{PFPD}. \quad (2)$$

The third and final step translates the confidence intervals generated by the ARIMA county "density" forecasts to the medium GMA county forecasts to produce confidence intervals around the CCM point total population forecast (GMAPOP):

$$\text{LLPOP} = \text{GMAPOP} - (\text{RLLPD} \times \text{GMAPOP}) \text{ and} \quad (3)$$

$$\text{UUPLP} = \text{GMAPOP} + (\text{RULPD} \times \text{GMAPOP}). \quad (4)$$

Appendix B contains the three different forecasts used in the ARIMA approach to measuring uncertainty in forecasts for Washington state and its 39 counties, and the Appendix B includes a numerical example of equations 1 through 4. Forecasts for 2030, 2040, and 2050 (10-to-30-year horizon lengths) are shown in three tables: Table B1 contains the ARIMA population density forecast and 95% confidence limits; Table B2 contains the GMA point forecast; and Table B3 includes the GMA point forecast and the translated 95% confidence limits.

Underlying the Espenshade-Tayman method is the idea that a sample is taken from a population of interest. In this case, the ARIMA results represent the sample, and the CCM forecasts represent the population. This interpretation is derived from the idea of a "superpopulation" (*Hartley – Sielken,* 1975; *Sampath,* 2005; *Swanson – Tayman,* 2012: 32–33). This concept can be traced back to *Deming and Stephan* (1941), who observed that even a complete census,

3) It is more common to use the term "forecast interval" or "prediction interval" in the context of forecasting, because a "confidence interval," strictly speaking, applies to a sample (*Swanson – Tayman,* 2014: 204). However, underlying our approach is the concept of a "super population," which describes a population as one sample from the infinity of populations (*Deming – Stephan,* 1941). Viewing a forecast as a sample leads us to choose "confidence interval" rather than forecast or prediction interval because it distinguishes the new approach from those discussed in the **Introduction**.

for scientific generalizations, describes a population that is but one of the infinities of populations that will result by chance from the same underlying social and economic causal systems. It is a theoretical concept that we use to simplify the application of statistical uncertainty to a population forecast that is considered a statistical model in this context.

ARIMA MODELS

Univariate ARIMA (Auto-Regressive Integrated Moving Average) time series models are the basis for the new approach for placing confidence intervals around CCM total population forecasts. The ARIMA model is described by *Box and Jenkins* (1976). It been used in the analysis and forecast of business, economic, and demographic variables (e.g., *Box et al.,* 2016; *Hyndman – Athanasopoulos,* 2024: Chapter 9; *Montgomery – Kulahci,* 2016). In recent years, ARIMA models have been developed using deep learning techniques (*Gridin,* 2022). Examples of its use in demographic forecasting include *McNown et al.* (1995); *Pflaumer* (1992); *Swanson* (2019); *Tayman, Smith, and Lin* (2007); and *Zakria and Muhammad* (2009). ARIMA models attempt to uncover the stochastic processes that generate a historical data series. The most general ARIMA model is usually written as ARIMA (p, d, q), where p is the order of the autoregression, d is the degree of differencing, and q is the order of the moving average. The autoregressive process has a memory in the sense that it is based on the correlation of each value of a variable with all preceding values. The moving average represents a "shock" to the system – an event with a substantial but short-lived impact on the time series pattern. The differencing process creates a stationary time series (i.e., one with a constant mean and variance over time). The d-value must be determined first because a stationary series is required to correctly identify the autoregressive and moving average processes.

We use the augmented Dickey-Fuller test (*Dickey – Fuller,* 1979) to identify the differencing required to achieve a stationary time series. The null hypothesis of this test is that a unit root is present in the time series, and the alternative hypothesis is that the time series is stationary. The patterns of the autocorrelation (ACF) and partial autocorrelation functions (PACF) are used to find the correct values for p and q (*Brockwell – Davis,* 2016: Chapter 3), and the autoregressive and moving average parameters have to be statistically significant. An adequate ARIMA model will have random residuals and the smallest possible values for p, d, or q. The Ljung-Box test (*Ljung – Box,* 1979) is used to evaluate the residuals of the estimated ARIMA model. The null hypothesis of this test is that the residuals are randomly distributed, and the alternative hypothesis is that the residuals are correlated with one another.

RESULTS

Table 1 presents selected statistics from the ARIMA models for each county and Washington state as a whole. The time series from 1960–2020 for all 39 counties and the state as a whole required differencing to become stationary. A significant Dickey-Fuller statistic ($p \leq 0.10$) based on first differences indicated a d = 1 in 29 counties and the state as a whole. The remaining ten counties required a second difference (difference of the first differences) in order to reject the null hypothesis of the presence of a unit root.[4] There is variation in the ARIMA parameters across counties. The most common specification was a model that contained only a first-order autoregressive term (19 counties). Nine counties contain only a first-order moving average term, while additional nine counties along with Washington state as a whole contain both first-order autoregressive and first-order moving average terms. Two counties have no autoregressive or moving average terms, referred to as random walk models. All counties and Washington state have

4) With the complete time series (1960–2020), Ferry County required a second difference to make the series stationary (Dickey Fuller p = 0.001). An ARIMA (1,2,1) for Ferry County showed illogical interval widths as the lower limit turned and stayed negative from 22- to 30-year forecast horizons. A graph comparing the first and second differences suggested that the non-stationarity in the time series occurred between 1960 and 1969. Table 1 shows that the restricted sample required only a first difference to achieve stationary (Dickey Fuller p = 0.036).

Ljung-Box p-values that exceed 0.10, indicating random residuals in all.

We conduct three analyses. First, we investigate the range of uncertainty in the county forecasts by analyzing county-specific half-widths (a half-width is defined as the width of the entire uncertainty interval divided by two). Second, to compare the examples discussed by *Yu et al.* (2023), Ferry, King, and Whitman counties, and Washington state as a whole, we compare their Bayesian-based intervals to our ARIMA-based intervals. We also include in the comparison the deterministic scenario-based intervals from the GMA forecasts. Finally, this analysis consists of a preliminary investigation of the effect of population size, growth rate, and horizon length on interval width.

Range of Uncertainty Across Counties

We begin by presenting the range of uncertainty in the county forecasts by analyzing half-widths (a half-width is defined as the width of the entire uncertainty interval divided by two. It represents the distance for the point estimate to either the upper or lower limit of the confidence interval) for 10-, 20-, and 30-year forecast horizons, as shown in Table 2. This Table also includes various summary measures of the half-width distribution across the counties for each of the three forecast horizons. As expected, the confidence intervals get wider with an increase in the forecast horizon for every county, as seen in the rise in the half-widths going from 10-year to 30-year forecast horizons. The percentage increases in the half-widths comparing the 10-year and 30-year forecast horizons range from 35.4% in Island County to 303.7% in Okanogan County; the average percent increase across counties is 111.1 (data not shown).

The summary measures of the half-widths tell a similar story: the average half-width increases from 10.5% in the 10-year forecast to 24.5% in the 30-year forecast, an increase of 133.3%. The half-width distributions are right skewed as the median half-widths are less than the mean half-widths in all forecast horizons. In the 10-year horizon, two counties have half-widths that exceed 20%. In the 20-year horizon, six counties have half-widths that exceed 35%, and in the 30-year horizon, six counties have half-widths that exceed 50%. The averages recomputed removing these cases are close to the median values reported in Table 2. Along with the average half-widths, the half-width variability across counties also increases with longer forecast horizons, with the coefficient of variation (abbreviated as "CV" in Table 2) rising from 48.3% in the 10-year forecast horizon to 75.5% in the 30-year forecast horizon. The direct relationship between the degree of forecast uncertainty and the length of the forecast horizon is well known. To our knowledge, this is the first study to empirically show that the variability of uncertainty also increases with the length of the forecast horizon.

A higher value of the d parameter causes wider ARIMA intervals and intervals that increase more rapidly with lengthening forecast horizons. For example, forecasts from an ARIMA model with first differences follow a linear trend, while forecasts from an ARIMA model with second differences will follow a quadratic trend (*Hyndman – Athanasopoulos,* 2024: Chapter 9; *Tayman – Smith – Lin,* 2007).

Table 1 **ARIMA Equations, Washington State and Counties**

		Coefficients				**Dickey Fuller Test p-Value**		
County	**ARIMA Specification**	**Auto Regressive**	**p-Value**	**Moving Average**	**p-Value**	**First Difference**	**Second Difference**	**Ljung-Box p-Value**
Adams	(1,1,1)	−0.430	0.001	−0.961	0.000	0.008		0.173
Asotin	(1,1,0)	0.338	0.004			0.009		0.159
Benton	(1,1,0)	0.814	0.000			0.005		0.546
Chelan	(0,2,1)			0.628	0.000	0.135	0.000	0.485
Clallam	(1,1,0)	0.560	0.000			0.055		0.857
Clark	(0,2,1)			0.277	0.031	0.202	0.001	0.857
Columbia	(1,1,0)	−0.209	0.100			0.006		0.890
Cowlitz	(1,1,0)	0.580	0.000			0.014		0.124
Douglas	(1,1,0)	0.347	0.004			0.002		0.285
Ferry[a]	(1,1,1)	0.857	0.000	0.558	0.004	0.036		0.120
Franklin	(1,2,0)	−0.350	0.004			0.424	0.001	0.841
Garfield	(0,1,1)			−0.441	0.000	0.003		0.168
Grant	(0,2,1)			0.776	0.000	0.171	0.000	0.211
Grays Harbor	(0,1,1)			−0.700	0.000	0.004		0.358
Island	(1,1,0)	0.224	0.077			0.024		0.766
Jefferson	(0,2,1)			0.550	0.000	0.143	0.001	0.112
King	(1,1,1)	0.558	0.000	−0.355	0.031	0.023		0.629
Kitsap	(1,1,0)	0.459	0.000			0.065		0.387
Kittitas	(1,2,1)	0.494	0.003	0.887	0.000	0.531	0.000	0.827
Klickitat	(0,2,1)			0.540	0.000	0.141	0.002	0.231
Lewis	(1,1,0)	0.436	0.000			0.006		0.739
Lincoln	(0,2,1)			0.584	0.000	0.120	0.000	0.111
Mason	(1,1,0)	0.733	0.000			0.087		0.365
Okanogan	(0,2,1)			0.487	0.000	0.132	0.001	0.897
Pacific	(1,1,0)	0.244	0.051			0.019		0.838
Pend Oreille	(1,1,1)	0.924	0.000	0.331	0.020	0.047		0.132
Pierce	(1,1,0)	0.341	0.005			0.002		0.492
San Juan	(1,1,0)	0.638	0.000			0.012		0.500
Skagit	(1,1,1)	0.902	0.000	0.390	0.010	0.105		0.608
Skamania	(0,1,0)					0.002		0.256
Snohomish	(1,1,0)	0.705	0.000			0.016		0.106
Spokane	(1,1,0)	704	0.000			0.097		0.925
Stevens	(1,1,0)	0.823	0.000			0.077		0.150
Thurston	(1,1,0)	0.584	0.000			0.004		0.219
Wahkiakum	(0,1,0)					0.004		0.113
Walla Walla	(1,1,1)	0.820	0.000	0.622	0.013	0.034		0.265
Whatcom	(1,1,0)	0.719	0.000			0.056		0.955
Whitman	(1,2,1)	−0.343	0.024	0.664	0.000	0.176	0.000	0.941
Yakima	(1,1,1)	0.932	0.000	0.450	0.001	0.099		0.722
Washington	(1,1,1)	0.687	0.000	−0.340	0.022	0.014		0.396

Note: a) The ARIMA model for Ferry County used an annual time series from 1970 to 2020, ten years shorter used for the other counties (see Footnote 4).

Table 2 **95% Half-Widths, ARIMA Alternative by Horizon Length Washington State and Counties, 2030–2050**[a)]

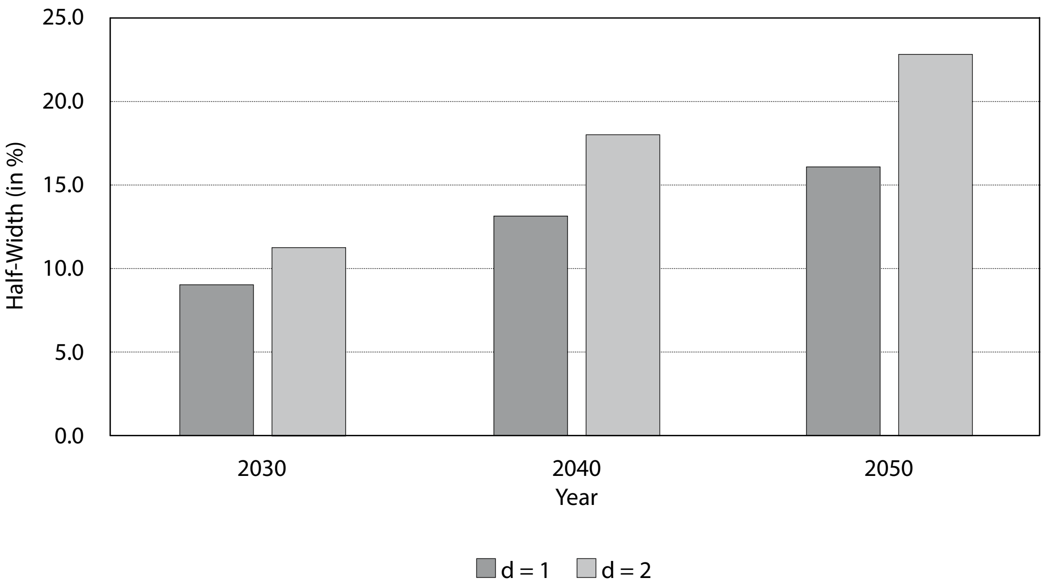
	Horizon Length		
County	**10-Year**	**20-year**	**30-year**
Adams	6.0%	8.0%	9.3%
Asotin	6.9%	9.0%	10.3%
Benton	13.0%	19.6%	23.3%
Chelan	11.5%	28.0%	43.9%
Clallam	7.8%	10.4%	11.8%
Clark	12.2%	24.3%	38.9%
Columbia	7.8%	12.5%	15.1%
Cowlitz	5.9%	8.2%	9.5%
Douglas	6.4%	8.2%	9.2%
Ferry	20.6%	29.7%	35.0%
Franklin	17.6%	37.7%	57.8%
Garfield	28.3%	41.4%	53.7%
Grant	12.6%	24.4%	36.7%
Grays Harbor	6.4%	8.6%	10.1%
Island	7.7%	9.5%	10.5%
Jefferson	16.0%	35.3%	56.1%
King	6.9%	9.3%	10.7%
Kitsap	8.6%	11.0%	12.2%
Kittitas	10.8%	19.1%	27.1%
Klickitat	17.4%	37.3%	58.2%
Lewis	5.4%	7.1%	8.1%
Lincoln	16.3%	38.2%	64.4%
Mason	8.7%	12.0%	13.7%
Okanogan	16.5%	39.2%	66.5%
Pacific	6.6%	8.7%	10.0%
Pend Oreille	17.0%	31.1%	40.9%
Pierce	5.2%	6.6%	7.4%
San Juan	10.7%	14.1%	15.6%
Skagit	10.2%	17.1%	21.6%
Skamania	6.5%	8.9%	9.2%
Snohomish	7.4%	10.1%	11.5%
Spokane	7.2%	10.3%	12.1%
Stevens	13.8%	21.4%	25.4%
Thurston	5.5%	7.2%	8.0%
Wahkiakum	11.0%	15.3%	17.9%
Walla Walla	6.6%	9.7%	11.7%
Whatcom	6.9%	9.7%	11.1%
Whitman	11.9%	25.7%	41.0%
Yakima	7.4%	13.7%	18.4%
Washington	5.5%	7.8%	8.9%

	Summary Statistics, Counties		
Mean	10.5%	17.9%	24.5%
Median	8.6%	12.5%	15.1%
Std. Dev.	5.1%	11.1%	18.5%
CV[b)]	48.3%	61.8%	75.5%

Notes: a) Calculated from Appendix Table B3 – Half Width = (High – Low) / 2 / Point Forecast × 100.
b) Std. Dev. / Mean × 100.

Turning to a verbal description of Figure 2, we first note that the average half-width for ARIMA models with first differences (the darker columns) is smaller than ARIMA models with second differences (the lighter columns) in all forecast years. Moreover, the gap between them increases with the lengthening of the forecast horizon. In 2030, the average half-widths are relatively similar (9.3% for d = 1 vs 10.9% for d = 2). By 2050, the gap has widened considerably (16.0 vs 23.2%). Put another way, the average of the half-width for ARIMA models with first differences increased by 173% compared to an increase of 214% for ARIMA models with second differences. In addition, the increase in the variability of the average half-width was much more significant for the ARIMA models with second differences. The percentage increase in the coefficients of variation from 2030 to 2050 is 122% (ARIMA with d = 1) and 211% (ARIMA with d = 2).

Figure 2 **Average Half-Width by Difference Parameter**



Comparison to Other Uncertainty Intervals for Selected Areas

As a means of evaluating the performance of our proposed method, we compare our forecast results to: (1) those that correspond with the results discussed by *Yu et al.* (2023), which are for Ferry, King and Whitman counties and the state of Washington as a whole; and (2) the judgmental intervals (the "low" and "high" values) that were placed by the OFM forecasters around their "middle" range projections (which were selected as the GMA forecasts) for all of the counties and the state as a whole. The comparisons are based on the "half-widths" of confidence intervals in Table 3. We focus on the "narrowness" of the half-widths because 95% confidence intervals may produce widths so wide as to be useless (*Swanson – Tayman,* 2014).

Table 3 **95% Half Width by Method and Horizon Length, Selected Counties and Washington State**[a)]

Ferry County				King County			
Horizon Length	**GMA[b)] Forecast**	**Bayes[c)] CCM**	**ARIMA[d)] Based**	**Horizon Length**	**GMA Forecast**	**Bayes CCM**	**ARIMA Based**
10 years	11.8%	9.7%	20.6%	10 years	10.6%	7.0%	6.9%
20 Years	21.6%	19.1%	29.7%	20 Years	15.2%	14.2%	9.3%
30 Years	32.0%	28.2%	35.0%	30 Years	19.3%	21.0%	10.7%

Whitman County				Washington State			
Horizon Length	**GMA Forecast**	**Bayes CCM**	**ARIMA Based**	**Horizon Length**	**GMA Forecast**	**Bayes CCM**	**ARIMA Base**
10 years	9.3%	4.9%	11.9%	10 years	9.6%	3.0%	5.5%
20 Years	11.8%	8.8%	25.7%	20 Years	13.4%	6.3%	7.8%
30 Years	14.3%	12.6%	41.0%	30 Years	16.6%	9.7%	8.9%

Note: a) Half Width = (High – Low) / 2 / Point Forecast × 100.
Sources: b) *Washington* (2022).
c) *Yu, et, al.* (2023).
d) Computed from Appendix Table B3.

Regarding Ferry County, Table 3 shows that the confidence intervals (half-widths) are widest for our approach (labeled as "ARIMA Based") for all horizon lengths, and those for the Bayes CCM approach (labeled as "Bayes CCM") are the narrowest for all three horizon lengths. The GMA judgmental half-width (labeled "GMA Forecast") falls between the other two approaches but tends to be closer to the Bayes CCM than ours. For all three approaches, the widths increase over time, per the expectation that forecast uncertainty increases as the horizon lengthens.

For King County at the 10-year horizon, the confidence intervals produced by our method are slightly narrower than the intervals reported both for the Bayesian method and those reported by OFM for the GMA forecast and substantially narrower at 20 and 30 years. The GMA intervals are somewhat narrower than the Bayes intervals at the 30-year horizon lengths but are not as narrow as our approach at any of the three horizon lengths. As was the case for Ferry County, the widths increase over time for all three approaches.

For Whitman County, the Bayes CCM approach produces the narrowest widths for all three horizon lengths. The GMA intervals are narrower than the intervals for our approach at all three horizon lengths, and those differences increase with the horizon length (2.6 percentage points for a 10-year horizon and 26.7 percentage points for a 30-year horizon).[5)] It should be noted, however, that *Yu et al.* (2023) held the age groups associated with college attendance constant in counties such as Whitman, where these populations significantly impact the county's overall age structure. Once again, the widths increase over time for all three approaches.

Considering Washington state as a whole, the intervals for the Bayes CCM approach are the narrowest for the 10- and 20-year horizon lengths, while our approach produces the narrowest interval for the 30-year horizon length. The GMA boundaries produce the widest intervals across all three horizon lengths by a sizable margin. Once again, the widths increase over time for all three approaches.

5) Whitman county's ARIMA model required second differences; and as expected, its interval width increased much faster than the other areas whose ARIMA models required first differences. From the 10-year to 30-year horizons, Whitman's half-width increased by 240% compared to 60% or 70% for the other two counties and Washington state.

These comparisons suggest that our approach produces uncertainty measures for county population forecasts similar to those generated by the Bayes CCM approach. Moreover, we find that all three approaches produce uncertainty intervals that are not so wide as to be useless, which is a point brought up by *Swanson and Tayman* (2016) in an earlier examination of forecast uncertainty.

Another example of the viability of our approach is Ferry County, which has the smallest population of the three counties. It has a 2020 population of 7,178 and a forecasted 2050 population of 6,986 (*Washington*, 2022). To see a 10-year horizon length half-width of 20.6 % and a half-width of 35.0% for the 30-year horizon length per our approach is not unexpected. Moreover, except for the 30-year horizon for Whitman County, all three methods across all three horizon lengths produce the widest uncertainty intervals in Ferry County, with the smallest population.

Impact of Population Size and Growth Rate on Interval Width

It has been established by ex-post evaluations that population size and growth rate affect forecast precision and bias (see *Smith – Tayman – Swanson,* 2013: 338–341 for a review of these findings). Consistent with these results and adding to them, we find that:

1. Forecast precision improves as population size increases, but this relationship weakens or disappears once the population reaches a certain size. However, population size has no predictable relationship with forecast bias.
2. Population growth rate affects both forecast precision and bias. Forecast precision is greatest for areas with small population changes, and declines as growth rates deviate in both a positive and negative direction from these low levels.
3. Bias is also strongly affected by differences in population growth rates. Areas losing population tend to be under-forecasted, whereas rapidly growing areas tend to be over-forecasted.

Now that we have shown our method produces intervals that are consistent with the idea that forecast interval width should increase temporally in conjunction with the increase in uncertainty expected as one looks further into the future, we turn to an examination of the effect that population size and growth rate have on forecast interval width. A review of the literature shows that such an examination has not yet been conducted with any specificity. To this end, we use a regression framework with the half-width as the dependent variable and population size and growth rate as the dependent variables, following the approach used in *Tayman, Smith, and Rayer* (2011). Separate models are estimated for 10- 20- and 30-year forecast horizons using single-variable regressions containing population size and growth rate and multiple regressions with both variables. Population size is measured at each forecast horizon, and growth rate is the percent change from 2020 to each forecast horizon.

We analyzed the functional form of the two independent variables at each forecast horizon length using graphs and the adjusted multiple coefficient of determination (adjR^2). The adjustment to the multiple coefficient of determination takes into account the complexity of a given regression model relative to the complexity of its input data (*Poston – Conde – Field,* 2024: 137–138). We determined the same functional form was appropriate for each horizon length. We illustrate this process using the 10-year forecast horizon, but the information for the 20- and 30-year horizons is available from us. Figure 3 shows the relationship between the natural log of population size (x axis) and half-width (y axis) for counties in Washington state at the 10-year forecast horizon. We use the natural log of the population to accommodate the skewed distribution of population size, which in 2030 ranges from 2,247 in Garfield County to 2,487,380 in King County. Also, the natural log of the population is more closely associated with the half-width than the unlogged population. In a single variable regression, the logged population has an adjR^2 of 0.272 compared to 0.067 for the unlogged value. As can be seen in Figure 3, the relationship tends to weaken around a population size of 36,000 (exp (ln (10.5)), suggesting the need for a squared term. However, adding a squared term slightly increases the adjR^2 from 0.213 to 0.231, which is not statistically significant.

Figure 4 shows the relationship between the decennial growth rate (x axis) and half-width (y axis) for counties in Washington state at the 10-year forecast horizon. The natural log of growth rate specification did not add to the explained variation of the half-width beyond the unlogged growth rate; both

Figure 3 **Relationship Between Ln(Population Size) and Half-Width, 2030**

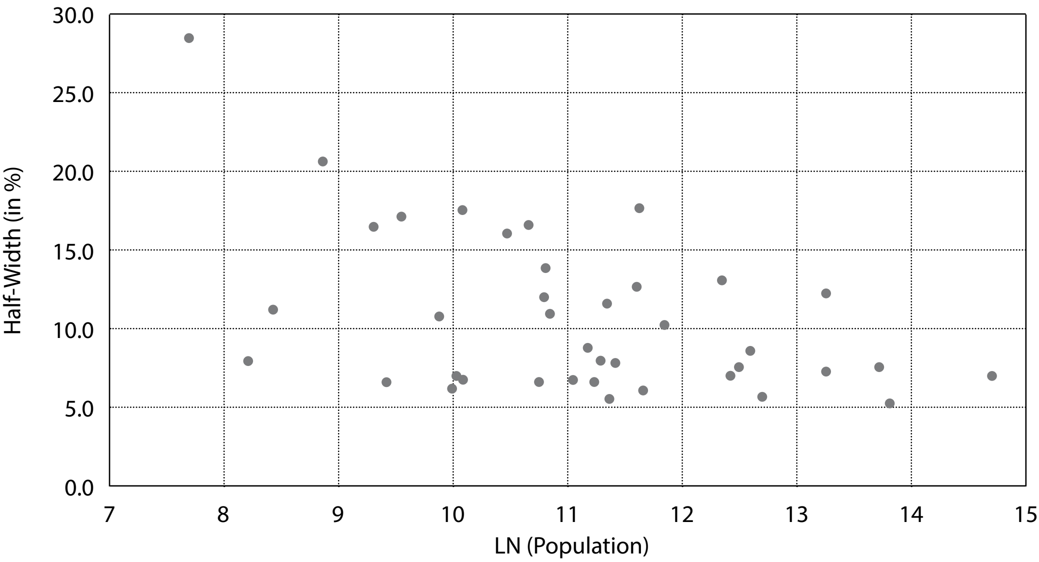
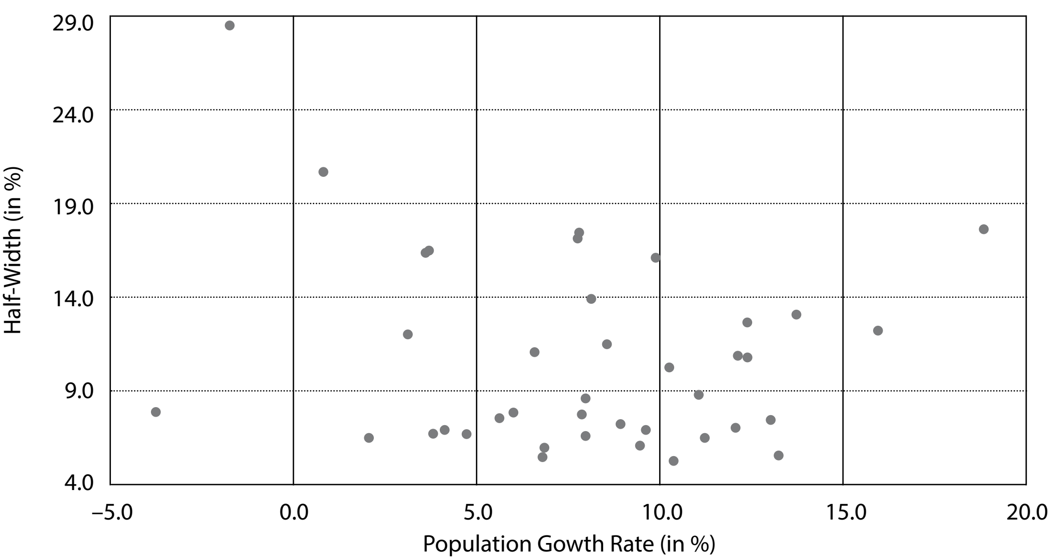


Figure 4 **Relationship Between Population Growth Rate and Half-Width, 2030**



specifications had an adjR^2 of 0.056. Growth rate and half-width rate have a u-shaped relationship with the half-width. Adding the squared term raises the adjR^2 substantially compared to an equation without it (0.006 vs 0.119).

The regression statistics are shown in Table 4, with the top two sections showing the adjR^2. The explanatory power of the models declines as the horizon length increases, similar to the results of *Tayman, Smith, and Rayer* (2011). However, contrary to the literature on population forecast error, population size has greater explanatory power than population growth rate in explaining the width of forecast intervals. The adjR^2 declines between −20% and −27% when the model includes only population size compared to the model that consists of both size and growth rate. The decline in adjR^2 is substantially greater when the model contains only growth rate, ranging from −65% in the 30-year forecast horizon to −55% in the 10-year forecast horizon.

Table 4 **Regression Models Predicting Half-With By Horizon Length Adjusted R^2 and Unstandarized Regression Coefficients**

	Adjusted R^2		
Model	*10-Year*	*20-Year*	*30-Year*
Ln (Size)	0.213	0.122	0.076
Growth Rate, Growth Rate^2	0.119	0.064	0.036
Ln (Size), Growth Rate, Growth Rate^2	0.267	0.166	0.104
	Reduction in R^2 From Size and Growth Rate Model		
Model	*10-Year*	*20-Year*	*30-Year*
Ln (Size)	−20.2%	−26.5%	−26.9%
Growth Rate, Growth Rate^2	−55.4%	−61.4%	−65.4%
	Unstandardized Regression Coefficients		
Variables	*10-Year*	*20-Year*	*30-Year*
Constant	0.291**	0.514**	−0.719**
Ln (Size)	−0.016**	−0.031**	−0.045*
Growth Rate	−0.539	−0.365	−0.255
Growth Rate^2	4.548*	2.051	1.254

Note: * $P < 0.10$, ** $P < 0.05$.

The last panel in Table 4 shows the regression coefficients for the model, containing both size and growth rate. The results for population size and growth rate are consistent across horizon lengths. The coefficients have the same signs, but their magnitudes vary less for population size than for population growth rate. The signs of the size coefficients are consistent with changes both in half-width and population size. They are also consistent with the parabolic relationship between growth rate and half-width. Half-width decreases as population losses moderate and increases as growth rates accelerate. Most coefficients are not statistically significant, partly due to the small sample size (39) and the number of variables in the regression equation. Four of the nine size and growth rate coefficients across all horizon years on the independent variables are significant at the 0.10 level. Three of the four significant coefficients are found in the population size variable.

DISCUSSION

The approach we propose can be linked directly not only to the CCM method but also to its algebraic equivalent, the CCR method. Unlike the approach found in *Swanson and Beck* (1994), neither the CCM nor the CCR approach is inherently conjoined with a method for generating statistical uncertainty. Thus, we believe this linkage represents a step forward on the path to generating probabilistic forecasts based on the fundamental population equation. In addition, this new approach is simpler than the methods described by *Cameron and Poot* (2011) and *Wilson* (2012) and far simpler than the Bayes CCM approach described by *Yu et al.* (2023). Moreover, it is neither opaque nor counter-intuitive, criticisms directed at Bayesian methods by *Goodwin* (2015). Notably, the ARIMA method is widely available in software packages used by state and local demographers and is more in line with their existing programming and other skills. They also have historical data that will support the construction of county ARIMA forecasts.

Analysis of the interval widths (as measured by the half-widths) is consistent with the expectation that probabilistic forecast intervals widen with increases in the forecast horizon. We found that the variability of the uncertainty across counties also increases as the forecast horizon increases. This represents a novel finding in that it does not appear in the literature. We examined the well-known non-linear relationships between population size and growth rate and forecast accuracy and bias using regression techniques with half-width as the dependent variable. Population size had a logarithmic relationship with half-width, and growth rate had a parabolic relationship (linear and squared terms). Interval width declined with increases in population size, but the interval width plateaued when counties reached a population of 36,000 or so. Interval widths were narrowest for slow-changing counties and increased as counties increased their rate of population decline or population increase. These findings are consistent with the relationships between population size and growth rate with forecast accuracy. What differs is that population size has a more substantial effect than population growth rate on interval width. The opposite occurs in the relationships with forecast accuracy. Keep in mind, however, these findings are based on a sample made up of the 39 counties in Washington state. They should be investigated in a larger sample of U.S. counties.

The strength of relationships between, on the one hand, population size and growth rate with, on the other, interval width was relatively weak and became weaker as the forecast horizon lengthened. Although not shown, the addition of a dummy variable representing counties with first and second-difference ARIMA models into the regression equations markedly raised the explained variance. Unlike the equations without the dummy variable, the explained variance increased as the forecast horizon lengthened. These results suggest that ARIMA model specification may be a more critical factor in explaining interval width than population size and growth rate. Future research should investigate and quantify the impact of these and other factors on interval widths from ARIMA models.

The approach we propose does not produce uncertainty intervals by age and gender, births, death, and migration, which are produced by the Bayes CCM approach described by *Yu et al.* (2023) and the CCR approach discussed by Swanson and Tayman (2014). Neither the Bayes CCM nor our approach, however, considers uncertainty in the input data themselves, a similarity also shared with the work by *Cameron and Poot* (2011), *Swanson and Tayman* (2014), and *Wilson* (2012). However, as *Yu et al.* (2023) implied, these are not likely to be among the most important sources of uncertainty for data in the United States and other countries where subnational population forecasts are routinely produced.

Regarding our approach not providing uncertainty intervals by age and gender, *Deming's* (1950: 127–134) "error propagation" was used to translate uncertainty in age group intervals found in the regression-based CCR forecasts reported by *Swanson and Tayman* (2014) to the total populations in question. In different forms, "error propagation" has been used by *Alho and Spencer* (2005), *Espenshade and Tayman* (1982), and *Hansen, Hurwitz, and Madow* (1953), among others. It may be possible to reverse-engineer error propagation and develop uncertainty measures by age and gender using our approach. It may be worthwhile to explore this possibility As an approximation, one could generate age uncertainty intervals by controlling the county "low" and "high" numbers in the 2017 GMA series to their corresponding 95%

lower and upper limits, respectively, of our proposed approach.

CONCLUSION

Smith, Tayman, and Swanson (2002: 373) opined that future research should focus increasingly on measuring uncertainty in population forecasts. Machine learning and AI may be significant in these endeavors (*Baker – Swanson – Tayman*, 2023). They noted that while such research may not directly improve forecast accuracy, it will enhance our understanding of the uncertainty inherent in population forecasts. They stated that this change would imply a shift from "population projections" to "population forecasts," a guideline we have followed in this paper.

In closing, we argue that the approach we propose and have described in this paper is well-suited for generating probabilistic subnational population forecasts in the United States and elsewhere where these forecasts are routinely produced. Because it can be applied to both the CCM and the CCR approaches, our method for producing forecast uncertainty information provides a path to a reasonable level of forecast accuracy as identified by *Swanson et al.* (2023). It also has the potential to optimize forecast utility, which as described in the **Introduction** is in accordance with the "triple constraint perspective" that underlies our approach. None of this is meant to imply that forecast uncertainty measures derived from ARIMA models using the Espenshade-Tayman method are more accurate than those generated from a Bayesian method. Rather, the findings herein suggest that our approach has a higher level of utility than a Bayesian approach while providing forecast intervals that are similar in width relative to both population size and forecast horizon length. As such, it offers a viable alternative to the Bayesian approach in that our results indicate that it produces similar measures of uncertainty, is simpler to implement, and, at this point in time, is likely to be more accessible to many of those who have been tasked to produce formal measures of uncertainty for their population forecasts.

Declarations and Acknowledgements

The data underlying this paper are secondary and available from the corresponding author.
This research was not funded by any agency.
The authors have no conflicting interests regarding this paper.
An ethics approval statement is not applicable because the data used are secondary, and no human subjects review was required.
A patient consent statement is not applicable.
A clinical trial registration statement is not applicable because no human subjects were involved.
A statement regarding permission to reproduce material from other sources is not applicable because no excerpts from copyrighted works owned by third parties are included.

References

- Alho, J. – Spencer, B. 2005. *Statistical demography and forecasting.* Springer B.V. Press: Dordrecht, Heidelberg, London, and New York.
- 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. – Alcantara, A. – Ruan, X. 2011. A stochastic version of the Brass PF Ratio Adjustment of age-specific fertility schedules. *PLOS One* 6(8): 23222. https://doi.org/10.1371/journal.pone.0023222.
- Baker, J. – Swanson, D. A. – Tayman, J. – Tedrow, L. 2017. *Cohort change ratios and their applications.* Springer B.V. Press: Dordrecht, Heidelberg, London, and New York. https://doi.org/10.1007/978-3-319-53745-0.
- Box, G. – Jenkins, G. 1976. *Time series analysis – Forecasting and control.* Holden-Day: San Francisco, CA.
- Box, G. – Jenkins, G. – Reinsel, G. – Ljung, G. 2016. *Time series analysis forecasting and control, 5th edition.* Wiley: Hoboken, NJ.
- Brockwell, P. J. – Davis, R. A. 2016. *Introduction to time series and forecasting, 3rd edition.* Springer Texts in Statistics: Switzerland. https://doi.org/10.1007/978-3-319-29854-2.
- Burch, T. 2018. *Model-based demography: Essays on integrating data, technique, and theory.* Demographic Research Monographs. Springer B.V. Press: Dordrecht, Heidelberg, London, and New York.
- Cameron, M. – Poot, J. 2011. Lessons from stochastic small-area population projections: The case of Waikato subregions in New Zealand. *Journal of Population Research,* 28: 245–265. https://doi.org/10.1007/s12546-011-9056-3.
- Canudas-Romo, V. – Land, K. – Yang, Y. – Yi, Z. 2008. Mathematical Demography. pp. 164–209. In Z. Yi (ed.). *Demography: Encyclopedia of Life Support Systems.* UNESCO-EOLSS.
- Deming, W. E. 1950. *Some theory of sampling.* New York, NY: Dover Publications.
- Deming, W. E. – Stephan, F. 1941. On the interpretation of censuses as samples. *Journal of the American Statistical Association,* 36(213): 45–49. https://doi.org/10.1080/01621459.1941.10502070.
- Dickey, D. A. – Fuller, W. A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association,* 74(366): 427–431. https://doi.org/10.2307/2286348.
- Espenshade, T. – Tayman, J. 1982. Confidence intervals for postcensal population estimates. *Demography,* 19(2): 191–210. https://doi.org/10.2307/2061190.
- George, M. V. – Smith, S. – Swanson, D. A. – Tayman, J. 2004. Population Projections. pp. 561–601. In J. Siegel – Swanson, D. (eds.). *The Methods and Materials of Demography, Condensed Edition, Revised.* 2004. Academic/Elsevier Press. https://doi.org/10.1016/B978-012641955-9/50055-2.
- Gridin, I. 2022. *Time series forecasting using deep learning.* BPB Publications: India.
- Goodwin, P. 2015. When simple alternatives to Bayes formula work well: Reducing the cognitive load when updating probability forecasts. *Journal of Business Research,* 68: 1686–1691. https://doi.org/10.1016/j.jbusres.2015.03.027.
- Hansen, N. – Hurwitz, W. – Madow, W. 1953. *Sample survey methods and theory, Volume I, methods and applications.* John Wiley and Sons: New York, NY (re-published in 1993).
- Hartley, H., – Sielken, R. Jr. 1975. A "super-population viewpoint" for finite population sampling. *Biometrics,* 31(2): 411–422. https://doi.org/10.2307/2529429.
- Hyndman, R. J. – Athanasopoulos, G. 2021. *Forecasting: principles and practice, 3rd edition,* OTexts: Melbourne, Australia. (OTexts.com/fpp3). Accessed on 9/17/2024.
- Land, K. 1986. Methods for national population forecasts: A review. *Journal of the American Statistical Association,* 81 (December): 888–901. https://doi.org/10.1080/01621459.1986.10478347.
- Ljung, G. M. – Box, G. E. 1978. On a measure of lack of fit in time series models. *Biometrika,* 65(2): 297–303. https://doi.org/10.1093/biomet/65.2.297.
- Montgomery, D. – Kulahci, M. 2016. *Introduction to time series analysis and forecasting, 2nd edition.* Hoboken, NJ: Wiley.
- McNown, R. – Rogers, A. – Little, J. 1995. Simplicity and complexity in extrapolative population forecasting models. *Mathematical Population Studies,* 5: 235–257. https://doi.org/10.1080/08898489509525404.
- Pflaumer, P. 1992. Forecasting U.S. population totals with the Box-Jenkins approach. *International Journal of Forecasting,* 8: 329–338. https://doi.org/10.1016/0169-2070(92)90051-A.

- Poston, D. – Conde, E. – Field, L. 2024. *Applied regression models in the social sciences.* Cambridge University Press. Cambridge, UK. https://doi.org/10.1017/9781108923071.
- Roe, L. – Swanson, D. A. – 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.
- Sampath, S. 2005. *Sampling theory and methods.* Alpha Science International Ltd. Harrow, England.
- Smith, S. – Tayman, J. – Swanson, D. A. 2001. *State and local population projections: Methodology and analysis.* Kluwer Academic /Plenum Press: New York. https://doi.org/10.1007/0-306-47372-0.
- Smith, S. – Tayman, J. – Swanson, D. A. 2013. *A practitioner's guide to state and local population projections.* Springer B.V. Press: Dordrecht, Heidelberg, London, and New York. https://doi.org/10.1007/978-94-007-7551-0_11.
- Stoto, M. 1983. The accuracy of population projections. *Journal of the American Statistical Association,* 78(381): 13–20. https://doi.org/10.1080/01621459.1983.10477916.
- Swanson, D. A. 2019. *Hopi tribal population forecast.* Report prepared for the Apache County Superior Court of the state of Arizona, in RE: The General Adjudication of all rights to use water in the Little Colorado River System and Source. CIVIL NO. 6417-203.
- Swanson, D. A. 1989. Confidence intervals for postcensal population estimates: A case study for local areas. *Survey Methodology,* 15(2): 271–280.
- Swanson, D. A. – Beck, D. 1994. A new short-term county level projection method. *Journal of Economic and Social Measurement,* 20: 25–50. https://doi.org/10.3233/JEM-1994-20102.
- Swanson, D. A. – Tayman, J. 2012. *Subnational population estimates.* Springer B.V. Press: Dordrecht, Heidelberg, London, and New York. https://doi.org/10.1007/978-90-481-8954-0.
- Swanson, D. A. – Tayman, J. 2014. Measuring uncertainty in population forecasts: A new approach. pp. 203–215 in Marco Marsili and Giorgia Capacci (eds.) *Proceedings of the 6th EUROSTAT/UNECE Work Session on Demographic Projections.* National Institute of Statistics, Rome, Italy.
- Swanson, D. A. – Burch, T. – Tedrow, L. 1996. What is applied demography? *Population Research and Policy Review,* 15 (December): 403–418. https://doi.org/10.1007/BF00125862.
- Swanson, D. A. – Bryan, T. – Hattendorf, M. – Comstock, K. – Starosta, L. – Schmidt, R. 2023. An example of combining expert judgment & small area projection methods: Forecasting for Water District needs. *Spatial Demography,* 11(8). https://doi.org/10.1007/s40980-023-00119-3.
- Tayman, J. – Swanson, D. A. 1996. On the utility of population forecasts. *Demography,* 33(4), 523–528. https://doi.org/10.2307/2061785.
- Tayman, J. – Smith, S. – Lin, J. 2007. Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models. Population Research and *Policy Review,* 26(3): 347–369. https://doi.org/10.1007/s11113-007-9034-9.
- Tayman, J. – Smith, S. – Rayer, S. 2011. Evaluating population forecast accuracy: A regression approach using county data. *Population Research and Policy Review,* 30(2): 235–262. https://doi.org/10.1007/s11113-010-9187-9.
- U.S. Census Bureau. 2023. State-produced population projections. https://www.census.gov/programs-surveys/popproj/about/related-sites.html.
- Washington 2020. 2019 *data book.* Forecasting and Research Division, Office of Financial Management, Washington state, Olympia, WA, https://ofm.wa.gov/sites/default/files/public/dataresearch/databook/pdf/databook.pdf.
- Washington 2022. Growth Management Act population projections for counties: 2020 to 2050 (last updated March 12, 2023): 2022 projections/population by age and sex, five-year age groups, medium series. Forecasting and Research Division, Office of Financial Management, Washington state, Olympia, WA. https://ofm.wa.gov/washington-data-research/population-demographics/population-forecasts-and-projections/growth-management-act-county-projections/growth-management-act-population-projections-counties-2020-2050.
- Washington 2024. Historical estimates of April 1 population and housing for the state, counties, and cities/intercensal estimate: 1960–2020. Forecasting and Research Division, Office of Financial Management, Washington state, Olympia, WA. https://ofm.wa.gov/washington-data-research/population-demographics/population-estimates/historical-estimates-april-1-population-and-housing-state-counties-and-cities.

- Wilson, T. 2012. Forecast accuracy and uncertainty of Australian Bureau of Statistics state and territory population projections. *International Journal of Population Research.* https://doi.org/10.1155/2012/419824.
- Wilson, T. 2005. Application of a probabilistic framework to New Zealand's official national population projections. *New Zealand Population Review,* 57(1): 51–76.
- Yu. C. – Ševčíková, H. – Raftery, A. – Curran, S. 2023. Probabilistic county-level population projections. *Demography,* 60(3): 915–937. https://doi.org/10.1215/00703370-10772782.
- Yusuf, F. – Martins, J. – Swanson, D. 2014. *Methods of Demographic Analysis.* Springer B.V. Press. https://doi.org/10.1007/978-94-007-6784-3.
- Zakria, M. – Muhammad, F. 2009. Forecasting the population of Pakistan using ARIMA models. *Pakistan Journal of Agricultural Sciences,* 46(3): 214–223.

DAVID A. SWANSON

is Distinguished Professor Emeritus, University of California Riverside and is affiliated with several population centers. Swanson has made contributions not only to statistical demography and formal demography, but also to applied demography, social demography, demographic methods, the collection of demographic data, and statistical methods. ScholarGPS places Swanson 56th among the world's 69 most highly ranked scholars in the field of demography. He earned his Ph.D. at the University of Hawai'i.

JEFF TAYMAN

specializes in the evaluation of demographic estimates and forecasts; the development of estimation and forecasting methods; the design and application of statistical methods; and the development of complex, integrated data systems for small geographic areas. Dr. Tayman has authored numerous articles and books. Among the most widely recognized books are: State and Local Population Projections: Methodology and Analysis (2001), Subnational Population Estimates (2012), A Practitioner's Guide to State and Local Population Projections (2013), and Cohort Change Ratios (2017). In 2023, Dr. Tayman retired as a Continuing Lecturer in the Department of Economics at the University of California San Diego.

APPENDIX

Appendix A: The Cohort Component Method of Population Projection

As its name suggests, the Cohort Component Method (CCM) requires the application of the components of population change – fertility, mortality, and migration to the age-gender structure at the projection's launch year (*George et al.,* 2004; *Smith – Tayman, – Swanson,* 2013: 155–182; *Yusuf – Martins – Swanson,* 2014: 231–253). There are three *components of change* in a population: mortality, fertility, and migration. The overall growth or decline of a population is determined by the interplay among these three components. The exact nature of this interplay can be formalized in the *basic demographic equation:*

$$P_l - P_b = B - D + IM - \text{OM} \qquad [1]$$

Where P_l is the population at the end of the time period; P_b is the population at the beginning of the time period; and *B, D, IM,* and *OM* are the number of births, deaths, in-migrants, and out-migrants during the time period, respectively. The difference between the number of births and the number of deaths is called *natural change (B – D);* it represents population growth coming from within the population itself. It may be either positive or negative, depending on whether births exceed deaths or deaths exceed births. The difference between the number of in-migrants and the number of out-migrants is called net *migration (IM – OM);* it represents population growth coming from the movement of people into and out of the area. It may be either positive or negative, depending on whether in-migrants exceed out-migrants or out-migrants exceed in-migrants.

The basic demographic equation can also be extended to apply to age groups, age-sex groups, and age-sex-race groups, as well as age-sex-ethnicity groups. This type of extension forms the logical basis of the and can be used to project a population into the future by age, age and sex, or by age, sex, and race. Once launched, these components (which are frequently modified as the projection moves into the future based on assumptions about their direction) are applied to the resulting age-gender structure at each cycle of the projection.

The Cohort Change Ratio Method of Population Projection

Unlike the CCM approach, its algebraic equivalent (*Baker et al.,* 2017: 251–252). The Cohort Change Ratio (CCR) method does not apply the separate components of population change to the age-sex structure at the launch year. Instead, it computes cohort change ratios (CCRs) using two counts of the age-structure in question, typically five or ten years apart, which directly capture mortality and migration. The fertility component uses a "child-adult ratio" from the most recent age structure data or a "child-woman ratio" for a projection by gender. It is well- suited for generating a projection of the population of the world, per the framework found in *Swanson et al.* (2023): (1) It corresponds to the dynamics by which a population moves forward in time; (2) there is information available relevant to these dynamics; (3) the time and resources needed to assemble relevant information and generate a projection is minimal; and (4) the information needed from the projection is generated by the CCR method.

The CCR method moves a population by age (and sex) from time t to time t+k using cohort-change ratios (CCRs) computed from data in the two most recent data points (e.g., censuses or estimates). It consists of two steps. The first uses existing data to develop CCRs, and the second applies the CCRs to the cohorts of the launch year population to move them into the future. The formula for the first step, the development of a CCR, is:

$$_nCCR_{x,i} = {_nP}_{x,i,t\,/\,nPx-k,i,t-k}, \qquad [2]$$

where

$_nP_{x,i,t}$ is the population aged x to $x+n$ in area i at the most recent census/estimate *(t)*,

$_nP_{x-k,i,t-k}$ is the population aged $x-k$ to $x-k+n$ in area i at the 2nd most recent

census/estimate *(t–k)*,

k is the number of years between the most recent census/estimate at time t

for area i and the census/estimate preceding it for area i at time $t-k$.

The basic formula for the second step, moving the cohorts of a population into the future, is:

$$ {}_nP_{x+k,i,t+k} = ({}_nCCR_{x,i}) \times ({}_nP_{x,i,t}), \qquad [3]$$

where

${}_nP_{x+k,i,t+k}$ is the population aged $x+k$ to $x+k+n$ in area i at time $t+k$

Given the nature of the CCRs, they cannot be calculated for the youngest age group (e.g., ages 0–4 if it is a five-year projection cycle; 0–9 if it is a ten-year projection cycle), because this cohort came into existence after the census/estimate data collected at time $t–k$. To project the youngest age group, we use the "Child-Adult Ratio" (CAR), where the number in the youngest age group at time t is divided by the number of adults at time t who are of childbearing age (e.g., 15–44). It does not require any data beyond what is available in the census/estimate sets of successive data.

The CAR equation for projecting the population aged 0–4 is:

$$\text{Population 0–4: } {}_5P_{0,t+}k = ({}_5P_{0,t} / {}_{30}P_{15,}t) \times ({}_{30}P_{15,t+k}), \qquad [4]$$

where

P is the population,

t is the year of the most recent census, and

$t+k$ is the estimation year.

Projections of the oldest open-ended age group differ slightly from the CCR projections for the age groups beyond age 10 up to the oldest open-ended age group. If, for example, the final closed age group is 80–84, with 85+ as the terminal open-ended age group, then calculations for the CCRi,x+ require the summation of the three oldest age groups to get the population age 75+ at time $t–k$:

$$ {}_\infty CCR_{75,i,t} = {}_\infty P_{85,i,t} / {}_\infty P_{75,i,t-k} \qquad [5]$$

The formula for estimating the population of 85+ of area i for the year t+k is:

$$ {}_\infty P_{85,i,t+k} = ({}_\infty CCR_{75,i,t}) \times ({}_\infty P_{75,i,t}). \qquad [6]$$

Appendix B: Forecasts Used in the ARIMA Approach to Measuring Uncertainty

Three different forecasts are used in the ARIMA approach to measure forecast uncertainty in the county and Washington state forecasts. Table B1 provides the population density forecasts for 2030, 2040, and 2050 by county and Washington state generated by the ARIMA method described in the paper. This Table also provides the land area of each county and Washington state. Table B2 provides the medium series of 2022 GMA (CCM) forecasts by county in 2030, 2040, and 2050, as well as for Washington state. These are the point forecasts used in translating the density confidence intervals to population confidence intervals. Finally, Table B3 shows the translation result. It provides the medium series GMA (CCM) forecasts by county for 2030, 2040, and 2050, as well as for Washington state, along with their 95% confidence intervals.

Forecast intervals for the 2030 Adams County population illustrate the ARIMA-based approach to measuring uncertainty. The lower limit, point forecast, and upper limit for the population density from Table B1 are 10.9, 11.6, and 12.3. respectively. Based on these densities, the relative distance between the limits and point forecast is derived by:

Lower Limit Distance (10.9 – 11.6) / 11.6 = –0.060345 and

Upper Limit Distance (12.3 – 11. 6) / 11.6 = 0.060345.

The Adams County 2030 Growth Management population "point" forecast is 22,565, as shown in Table B2. The confidence intervals around the 2030 population point forecast shown in Table B3 are derived by:

Lower Limit Population 22,565 + (–0.060345 × 22,565) = 21,203 and

Lower Limit Population 22,565 + (0.060345 × 22,565) = 23,927.

Table B1 **ARIMA Population Density Forecasts, Washington State and Counties, 2030–2050**[a)]

		2030			2040			2050		
County	**Land Area**[b)]	**LL95%**	**Point**	**UL95%**	**LL95%**	**Point**	**UL95%**	**LL95%**	**Point**	**UL95%**
Adams	1,924	10.9	11.6	12.3	11.5	12.5	13.5	12.1	13.4	14.6
Asotin	636	35.2	37.8	40.4	37.0	40.6	44.3	39.0	43.5	48.0
Benton	1,703	121.6	139.7	157.8	124.5	154.9	185.3	130.2	169.7	209.2
Chelan	2,921	26.2	30.1	33.1	23.9	33.2	42.5	20.3	36.2	52.1
Clallam	1,745	45.1	48.9	52.7	48.1	53.7	59.3	51.6	58.5	65.4
Clark	628	868.3	988.9	1,109.4	876.3	1,157.4	1,438.5	833.0	1,363.4	1,893.8
Columbia	869	4.1	4.5	4.8	3.8	4.4	4.9	3.6	4.3	4.9
Cowlitz	1,139	99.1	105.3	111.6	103.6	112.8	122.0	108.8	120.2	131.6
Douglas	1,820	24.7	26.4	28.1	26.8	29.2	31.6	29.0	32.0	34.9
Ferry	2,210	2.7	3.4	4.1	2.6	3.7	4.8	2.6	4.0	5.4
Franklin	1,242	77.7	94.3	110.9	69.0	110.7	152.4	53.6	127.1	200.6
Garfield	710	2.2	3.0	3.9	1.7	2.9	4.1	1.2	2.7	4.1
Grant	2,676	36.5	41.8	47.0	35.1	46.5	57.8	32.4	51.2	70.0
Grays Harbor	1,917	38.5	41.2	43.8	39.2	42.9	46.6	40.1	44.6	49.1
Island	209	439.3	475.8	512.8	484.6	535.6	586.7	533.2	595.4	657.7
Jefferson	1,808	16.8	20.0	23.2	14.1	21.8	29.5	10.3	23.6	36.8
King	2,126	1,101.4	1,182.8	1,264.1	1,163.0	1,282.5	1,402.0	1,234.0	1,382.2	1,530.4
Kitsap	396	718.9	786.2	853.5	780.8	877.1	973.3	849.6	968.0	1,086.3
Kittitas	2,297	21.1	23.7	26.2	21.7	26.9	32.0	21.9	30.1	38.2
Klickitat	1,871	11.4	13.8	16.2	9.6	15.4	21.1	7.1	17.0	26.9
Lewis	2,408	35.2	37.2	39.2	37.3	40.1	43.0	39.6	43.1	46.6
Lincoln	2,313	4.1	4.9	5.7	3.2	5.1	7.1	2.0	5.2	8.7
Mason	961	70.5	77.2	83.9	76.4	86.8	97.2	83.2	96.4	109.6
Okanogan	5,273	6.8	8.2	9.5	5.1	8.3	11.6	2.9	8.5	14.2
Pacific	974	23.9	25.6	27.3	24.8	27.1	29.5	25.8	28.6	31.5
Pend Oreille	1,400	8.5	10.3	12.0	7.7	11.1	14.6	7.1	12.1	17.0
Pierce	1,675	579.3	610.9	642.4	626.5	671.0	715.5	676.7	731.2	785.6
San Juan	175	105.3	117.9	130.6	115.1	133.9	152.8	126.5	150.0	173.4
Skagit	1,735	75.4	83.9	92.5	77.0	92.9	108.8	79.8	101.7	123.7
Skamania	1,658	7.2	7.7	8.2	7.7	8.4	9.2	8.3	9.2	10.0
Snohomish	2,090	419.6	453.2	486.7	456.5	508.0	559.5	498.1	562.8	627.5
Spokane	1,764	311.5	335.6	359.8	323.7	360.8	397.9	339.1	385.7	432.7
Stevens	2,477	17.8	20.7	23.5	18.0	22.9	27.8	18.8	25.2	31.6
Thurston	727	439.8	465.4	490.9	485.9	523.4	560.9	535.0	581.4	627.9
Wahkiakum	264	15.3	17.2	19.1	15.0	17.7	20.4	14.9	18.2	21.4
Walla Walla	1,271	48.9	52.3	55.8	50.0	55.4	60.8	51.6	58.5	65.3
Whatcom	2,120	112.0	120.4	128.7	120.7	133.6	146.5	130.6	146.9	163.2
Whitman	2,159	21.0	23.9	26.7	18.9	25.5	32.0	16.0	27.1	38.2
Yakima	4,296	59.2	63.9	68.7	59.2	68.6	78.0	60.0	73.5	87.1
Washington	**66,589**	**122.6**	**129.8**	**137.0**	**131.3**	**142.4**	**153.4**	**141.0**	**154.9**	**168.7**

Note: a) Population per square mile.
Source: b) Land area in square miles (Washington, 2020); Also see discussion in text.

Table B2 **2022 GMA Medium Forecasts, Washington State and Counties, 2030–2050**

County	2030	2040	2050
Adams	22,565	24,387	26,100
Asotin	23,214	23,815	24,111
Benton	235,177	262,587	288,887
Chelan	85,889	91,914	97,195
Clallam	81,791	85,374	87,800
Clark	583,307	660,653	735,724
Columbia	3,806	3,625	3,366
Cowlitz	118,309	125,320	130,993
Douglas	47,750	52,256	56,461
Ferry	7,239	7,169	6,986
Franklin	114,907	132,930	150,970
Garfield	2,247	2,172	2,061
Grant	111,367	123,116	134,321
Grays Harbor	77,203	77,614	76,892
Island	93,670	99,870	105,250
Jefferson	36,226	39,170	41,719
King	2,487,380	2,690,851	2,879,176
Kitsap	297,608	317,694	335,268
Kittitas	52,091	57,521	62,643
Klickitat	24,511	26,059	27,376
Lewis	87,746	92,313	95,871
Lincoln	11,270	11,459	11,496
Mason	72,981	79,792	85,947
Okanogan	43,676	44,660	45,101
Pacific	24,475	25,033	25,183
Pend Oreille	14,442	15,311	16,009
Pierce	1,015,395	1,104,062	1,186,146
San Juan	19,986	22,046	23,957
Skagit	142,805	155,142	166,281
Skamania	12,529	13,322	14,006
Snohomish	935,370	1,039,254	1,138,649
Spokane	587,377	630,994	669,671
Stevens	50,215	53,502	56,278
Thurston	333,783	371,542	407,392
Wahkiakum	4,713	4,925	5,070
Walla Walla	64,977	66,695	67,645
Whatcom	254,158	280,275	304,836
Whitman	49,489	50,698	51,459
Yakima	271,120	283,351	293,279
Washington	8,502,764	9,248,473	9,937,575

Source: Washington, 2022

Table B3 **ARIMA 95% Intervals Applied to the GMA Medium Forecasts, Washington State and Counties, 2030–2050**

	2030			2040			2050		
County	**LL95%**	**Point**	**UL95%**	**LL95%**	**Point**	**UL95%**	**LL95%**	**Point**	**UL95%**
Adams	21,203	22,565	23,927	22,436	24,387	26,338	23,568	26,100	28,437
Asotin	21,617	23,214	24,811	21,703	23,815	25,985	21,617	24,111	26,605
Benton	204,707	235,177	265,647	211,053	262,587	314,121	221,645	288,887	356,129
Chelan	74,761	85,889	94,449	66,167	91,914	117,661	54,504	97,195	139,886
Clallam	75,435	81,791	88,147	76,471	85,374	94,277	77,444	87,800	98,156
Clark	512,171	583,307	654,384	500,199	660,653	821,107	449,507	735,724	1,021,941
Columbia	3,468	3,806	4,060	3,131	3,625	4,037	2,818	3,366	3,836
Cowlitz	111,343	118,309	125,387	115,099	125,320	135,541	118,569	130,993	143,417
Douglas	44,675	47,750	50,825	47,961	52,256	56,551	51,168	56,461	61,578
Ferry	5,749	7,239	8,729	5,038	7,169	9,300	4,541	6,986	9,431
Franklin	94,679	114,907	135,135	82,856	132,930	183,004	63,666	150,970	238,274
Garfield	1,648	2,247	2,921	1,273	2,172	3,071	916	2,061	3,130
Grant	97,246	111,367	125,221	92,933	123,116	153,035	85,000	134,321	183,642
Grays Harbor	72,144	77,203	82,075	70,920	77,614	84,308	69,134	76,892	84,650
Island	86,484	93,670	100,954	90,360	99,870	109,398	94,255	105,250	116,263
Jefferson	30,430	36,226	42,022	25,335	39,170	53,005	18,208	41,719	65,053
King	2,316,199	2,487,380	2,658,351	2,440,125	2,690,851	2,941,577	2,570,470	2,879,176	3,187,882
Kitsap	272,132	297,608	323,084	282,813	317,694	352,539	294,260	335,268	376,241
Kittitas	46,376	52,091	57,586	46,402	57,521	68,426	45,577	62,643	79,500
Klickitat	20,248	24,511	28,774	16,245	26,059	35,704	11,434	27,376	43,318
Lewis	83,028	87,746	92,464	85,867	92,313	98,989	88,086	95,871	103,656
Lincoln	9,430	11,270	13,110	7,190	11,459	15,953	4,422	11,496	19,234
Mason	66,647	72,981	79,315	70,232	79,792	89,352	74,178	85,947	97,716
Okanogan	36,219	43,676	50,600	27,442	44,660	62,416	15,387	45,101	75,345
Pacific	22,850	24,475	26,100	22,908	25,033	27,250	22,718	25,183	27,737
Pend Oreille	11,918	14,442	16,826	10,621	15,311	20,139	9,394	16,009	22,492
Pierce	962,872	1,015,395	1,067,752	1,030,842	1,104,062	1,177,282	1,097,737	1,186,146	1,274,393
San Juan	17,850	19,986	22,139	18,951	22,046	25,158	20,204	23,957	27,694
Skagit	128,337	142,805	157,443	128,589	155,142	181,695	130,474	166,281	202,251
Skamania	11,715	12,529	13,343	12,212	13,322	14,591	12,636	14,006	15,224
Snohomish	866,022	935,370	1,004,511	933,897	1,039,254	1,144,611	1,007,749	1,138,649	1,269,549
Spokane	545,196	587,377	629,733	566,111	630,994	695,877	588,762	669,671	751,275
Stevens	43,180	50,215	57,007	42,054	53,502	64,950	41,985	56,278	70,571
Thurston	315,423	333,783	352,071	344,922	371,542	398,162	374,879	407,392	439,975
Wahkiakum	4,192	4,713	5,234	4,174	4,925	5,676	4,151	5,070	5,961
Walla Walla	60,753	64,977	69,325	60,194	66,695	73,196	59,666	67,645	75,508
Whatcom	236,426	254,158	271,679	253,213	280,275	307,337	271,011	304,836	338,661
Whitman	43,484	49,489	55,287	37,576	50,698	63,621	30,382	51,459	72,536
Yakima	251,178	271,120	291,486	244,524	283,351	322,178	239,411	293,279	347,546
Washington	**8,031,116**	**8,502,764**	**8,974,412**	**8,527,560**	**9,248,473**	**9,962,892**	**9,045,824**	**9,937,575**	**10,822,911**

Sources: Data from Tables B1 and B2, calculations of intervals by authors.