Demography, Vol. 25, No. 3, August 1988

Stability Over Time in the Distribution of Population Forecast Errors

Stanley K. Smith and Terry Sincich
College of Business Administration, University of Florida, Gainesville, Florida 32611

A number of studies in recent years have investigated empirical approaches to the production of confidence intervals for population projections. The critical assumption underlying these approaches is that the distribution of forecast errors remains stable over time. In this article, we evaluate this assumption by making population projections for states for a number of time periods during the 20th century, comparing these projections with census enumerations to determine forecast errors, and analyzing the stability of the resulting error distributions over time. These data are then used to construct and test empirical confidence limits. We find that in this sample the distribution of absolute percentage errors remained relatively stable over time and data on past forecast errors provided very useful predictions of future forecast errors.

A number of recent studies have investigated empirical approaches to the production and use of confidence intervals for population projections. Some have used time series models in which historical population data were fit to autoregressive or moving average processes and future values were made to depend on a weighted average of past values and a random error term (e.g., Lee, 1974; Saboia, 1974; Voss et al., 1981). Others have focused directly on the errors of projections made in the past (e.g., Keyfitz, 1981; Smith, 1987; Stoto, 1983). Under either of these approaches, the critical assumption underlying the construction and use of confidence intervals for population projections is that the distribution of forecast errors remains stable over time.

Is that a reasonable assumption? To date, little research has been directed toward answering this question. In the present study, we analyze the distribution of population forecast errors for states for a number of different time periods during the 20th century. Our objectives are twofold: to determine the extent to which these distributions have remained stable over time and to evaluate the validity of using data on the distribution of past forecast errors to predict the distribution of future forecast errors. The next section describes the data and techniques used to make the population projections analyzed in this study. The third section discusses the characteristics of forecast errors and the stability of those errors over time. The fourth section uses data on the distribution of past forecast errors to construct empirical confidence limits and tests their performance in predicting the range of future forecast errors. The fifth section draws several conclusions regarding stability in the distribution of forecast errors over time and the potential usefulness of developing empirical confidence limits for population projections.

Demographers often draw a distinction between the terms "projection" and "forecast." A projection is typically defined as the numerical outcome of a specific set of assumptions regarding future trends, whereas a forecast indicates the specific projection that the author believes is most likely to provide an accurate prediction of future population (e.g., Irwin, 1977; Isserman and Fisher, 1984; Keyfitz, 1972). In this article, projection refers to the future population implied by a particular technique and data set, and forecast accuracy refers to the percentage difference between a projection and the census-enumerated population for the same year. In other words, projections are treated as if they were indeed forecasts of future population.

Data and Projection Techniques

The data used in this study were the final population counts from each decennial census from 1900 to 1980, for each of the 50 states in the United States (U.S. Bureau of the Census, 1983). These data refer to total population only; no analysis was performed on the age, sex, or race distribution of the population. The following terminology, based in part on Cohen (1986), is used to describe population projections:

1. Base year—the year of the earliest observed population size used to make a projection.
2. Launch year—the year of the latest observed population size used to make a projection.
3. Target year—the year for which population size is projected.
4. Base period—the interval between the base year and the launch year.
5. Projection horizon—the interval between the launch year and the target year.

For example, if data from 1900 and 1910 were used to project population size in 1920, then 1900 would be the base year; 1910, the launch year; 1920, the target year; 1900–1910, the base period; and 1910–1920, the projection horizon.

Four primary projection techniques were used. First was linear extrapolation (Line), which assumes that a population will increase (decrease) by the same number of persons in each future year as the average annual increase (decrease) during the base period:

$$P_t = P_l + x/y(P_l - P_b), \tag{1}$$

where P_t = state population projection for target year, P_l = state population in launch year, P_b = state population in base year, x = number of years in projection horizon, and y = number of years in base period.

The second technique was exponential extrapolation (Expo), which assumes that a population will grow (decline) at the same annual rate in each future year as during the base period:

$$P_t = P_l \exp(rx), \tag{2}$$

where r = average annual growth rate during base period.

In the third and fourth techniques, state population data were expressed as shares of national population data. Under the shift-share technique (Shift), state shares of national population were calculated for the base year and launch year and were extrapolated into the future by assuming that the average annual absolute change in each state's share of national population observed during the base period would continue throughout the projection horizon. The extrapolated state shares were then applied to an independent projection of national population to provide state population projections:

$$P_t = P_{jt}\,[P_l/P_{jl} + x/y\,(P_l/P_{jl} - P_b/P_{jb})], \tag{3}$$

where P_{jt} = national population projection for target year, P_{jl} = national population in launch year, and P_{jb} = national population in base year.

Under the share-of-growth technique (Share), state shares of national population growth during the base period were calculated. Projections were made by assuming that these shares would be the same throughout the projection horizon as they were during the base period:

$$P_t = P_l + [(P_l - P_b)/(P_{jl} - P_{jb})] (P_{jt} - P_{jl}). \quad (4)$$

The Shift and Share techniques each require an independent projection of the national population. Since no widely accepted set of national projections was produced in the early part of this century, a new set had to be constructed. This was done by applying the Line and Expo techniques to the United States population and using an average of the results as a national projection.

It should be noted that the Shift and Share techniques have two sources of error: one caused by errors in projecting state shares of the national population and the other caused by errors in the national projections. The latter source of error is minor. An alternative set of projections from Shift and Share was made using census enumerations of the national population for the target years instead of projections. The results from this alternative set of state projections differed only slightly from those reported in this article. National projections rather than census data were used in this analysis because employing the Shift and Share techniques for population projections requires the use of projections rather than census counts of the larger area population.

A fifth technique was evaluated as well. Projections from this technique (Ave) were simply the averages of the projections produced by the four primary techniques.

Simple projection techniques have a long history, having been used by such notable persons as Benjamin Franklin, Thomas Jefferson, and Abraham Lincoln (Dorn, 1950:315). They are still frequently used for local population projections (Federal–State Cooperative Program, 1984) but are no longer commonly used for projections of state and national populations, having been replaced by more sophisticated cohort-component and economic-demographic techniques. These more sophisticated techniques, however, were developed and used for state-level projections only within the last several decades (e.g., U.S. Bureau of the Census, 1957; U.S. Bureau of Economic Analysis, 1974). Projections from these techniques are therefore not available for most decades of this century. Creating such projections specifically for this study would have been prohibitively expensive or even impossible, given the lack of relevant data. Consequently, it was necessary to base the present analysis on the simple projection techniques described above.

The simplicity of these techniques does not negate their usefulness, however. A number of studies have concluded that simple extrapolation techniques produce short- to medium-term forecasts of total population that are at least as accurate as those produced by more sophisticated techniques (e.g., Ascher, 1978; Greenberg, 1972; Kale et al., 1981; Murdock et al., 1984; Siegel, 1953, 1972; Smith, 1984; White, 1954). Furthermore, the more sophisticated techniques themselves are typically based on extrapolations of one type or another (e.g., migration rates, birth rates, and survival rates for cohort-component projections; employment trends for economic-based projections). The functional forms of these extrapolations are often similar to those of the simple techniques. If applied to the same base periods, then, the projections from the more sophisticated techniques would most likely be similar to those analyzed in this study (Smith, 1987).

The advantages of using simple projection techniques for the present analysis are that they require very few base data, they can be applied at low cost, and they can be applied retrospectively to produce a large number of consistent projections that are comparable over time. We believe these simple techniques provide a useful vehicle for testing stability in the distribution of population forecast errors over time.

Analysis of Errors

Projections for each of the 50 states were made for 10- and 20-year horizons, using 10-year base periods and the five techniques described previously. Projections were made from every base period since 1900, yielding seven sets of 10-year projections and six sets of 20-year projections. These projections were then compared with decennial census counts for each target year. The resulting differences are called forecast errors, although they might have been caused partly by errors in census enumeration as well as by errors in the forecasts themselves. Throughout this article the term "error" refers to percentage error rather than to absolute numerical error.

General Characteristics

Two measures were used to provide a general description of forecast error characteristics. Mean absolute percentage error (MAPE) is the average percentage error when the direction of error is ignored. This provides a measure of forecast accuracy. Mean algebraic percentage error (MALPE) is the average percentage error when the direction of error is accounted for. This provides a measure of bias: a positive error indicates a tendency for projections to be too high and a negative error indicates a tendency for projections to be too low.

To provide an overall picture of forecast error characteristics, errors were aggregated from all of the sets of 10- and 20-year projections, providing a pooled sample of errors from 350 10-year projections and 300 20-year projections. Tables 1 and 2 summarize these errors

Table 1. MAPE and MALPE by Technique, Population Size, and Length of Projection Horizon

Technique	Population size*	MAPE, projection horizon		MALPE, projection horizon	
		10 years	20 years	10 years	20 years
Line	<1 million	9.5	17.3	−1.2	−3.3
	1–3 million	6.0	11.9	−1.4	−2.3
	>3 million	5.5	9.1	−0.2	−1.4
	Total	7.0	13.0	−1.0	−2.4
Expo	<1 million	11.6	26.8	3.7	12.2
	1–3 million	7.4	16.3	1.2	5.9
	>3 million	6.6	13.5	2.3	6.3
	Total	8.5	19.1	2.3	8.1
Shift	<1 million	9.9	18.3	−0.2	−0.5
	1–3 million	6.2	12.3	−0.7	−0.1
	>3 million	5.8	9.8	0.7	1.2
	Total	7.3	13.7	−0.1	0.1
Share	<1 million	10.6	21.6	0.0	0.1
	1–3 million	7.0	14.3	−1.3	−2.0
	>3 million	6.0	11.1	0.7	1.1
	Total	7.9	15.9	−0.2	−0.4
Ave	<1 million	10.3	20.3	0.6	2.1
	1–3 million	6.6	13.4	−0.6	0.4
	>3 million	5.9	10.5	0.9	1.8
	Total	7.6	15.0	0.2	1.4

* Population size in launch year.

Table 2. MAPE and MALPE by Technique, Rate of Growth, and Length of Projection Horizon

Technique	Rate of growth*	MAPE, projection horizon		MALPE, projection horizon	
		10 years	20 years	10 years	20 years
Line	<10%	5.7	10.8	−3.0	−6.2
	10–20%	5.6	10.4	−0.5	−2.2
	>20%	10.9	19.2	1.3	2.6
	Total	7.0	13.0	−1.0	−2.4
Expo	<10%	5.6	10.6	−2.7	−5.2
	10–20%	5.9	11.4	1.1	2.4
	>20%	16.4	41.1	11.5	34.5
	Total	8.5	19.1	2.3	8.1
Shift	<10%	5.7	10.9	−2.8	−5.4
	10–20%	5.8	10.9	0.3	0.3
	>20%	11.7	21.2	3.2	7.7
	Total	7.3	13.7	−0.1	0.1
Share	<10%	6.5	14.2	−4.7	−11.2
	10–20%	5.6	10.6	0.1	−0.7
	>20%	13.0	25.2	5.9	15.1
	Total	7.9	15.9	−0.2	−0.4
Ave	<10%	5.8	11.4	−3.3	−7.0
	10–20%	5.7	10.8	0.2	−0.1
	>20%	12.8	25.5	5.5	15.0
	Total	7.6	15.0	0.2	1.4

* Rate of growth during base period.

for each technique, with states divided according to population size in the launch year (Table 1) and rate of population growth during the base period (Table 2).

A number of distinct patterns can be seen in these tables. Forecast accuracy (a) declined as the length of the projection horizon increased, (b) increased as the size of the base population increased, and (c) declined as the rate of population growth during the base period increased. These patterns were found for all five techniques and for both 10- and 20-year projections. Similar patterns were found in many other studies (e.g., Irwin, 1977; Isserman, 1977; Schmitt and Crosetti, 1951, 1953; Smith, 1987; Stoto, 1983; White, 1954).

There was no evidence of any overall bias in these projections. The total sample MALPE was slightly negative for Line, slightly positive for Expo, and virtually zero for Shift and Share. In no instance were total sample MALPEs significantly different than zero. There was no evidence of a consistent relationship between population size and bias (Table 1), but there was strong evidence of one between growth rates and bias (Table 2). For every technique in both the 10- and 20-year projections, MALPEs increased as the rate of growth during the base period increased. Differences in MALPEs by rate of growth were frequently very large. A positive relationship between bias and population growth during the base period was previously noted for projections of counties in the United States (Smith, 1987).

A number of studies have concluded that the choice of projection technique generally has little impact on the overall forecast accuracy of short- to medium-term population projections, once the base period and projection horizon have been fixed (e.g., Ascher, 1978; Kale et al., 1981; Smith, 1984; White, 1954). This study supports that conclusion. Differences in forecast accuracy among alternative techniques were found to be small in most instances. The overall MAPEs for the 10-year projections were between 7.0 percent and 8.5 percent for all five techniques, whereas the 20-year projections displayed a wider range of 13.0 percent to 19.1 percent. Expo had the largest errors, particularly for the 20-year projections and for states with populations of less than 1 million or growth rates of greater than 20 percent. Other than these results for Expo, however, the errors for the different techniques were similar. Furthermore, the patterns relating forecast errors to size of place, rate of growth, and length of projection horizon were much the same for all five techniques.

Stability Over Time

The pooled sample data summarized in Tables 1 and 2 provide an overall picture of the error characteristics for state population projections during the 20th century. The results regarding accuracy, bias, and the effects of population size, growth rate, and length of projection horizon are consistent with the findings of many other studies. These results tell us nothing, however, about trends in forecast accuracy and bias over time. What patterns can be seen when each set of projections is viewed individually?

Table 3 shows MAPEs and MALPEs for 10-year projections for each technique and target year from 1920 to 1980. Several conclusions can be drawn from this table. First, the largest MAPEs were for the first set of projections (1920). This was true for all five techniques, most notably for Expo.[1] Second, there were no persistent biases in these projections. For each technique MALPEs were positive for about half of the target years and negative for about half. Third, for target years after 1920, there were no apparent trends in accuracy or bias. MAPEs and MALPEs did not become systematically larger or smaller over time but, rather, fluctuated up and down. Finally, after 1920 MAPEs and MALPEs did not differ dramatically by target year or technique. On the contrary, they typically fell within a fairly narrow range. Similar results regarding MAPEs and MALPEs were found in an analysis of 20-year projections for target years 1930–1980 (not shown here).

Table 3. MAPE and MALPE for 10-Year Projections, by Technique and Target Year

Target year	Technique				
	Line	Expo	Shift	Share	Ave
			MAPE		
1920	8.4	15.6	9.5	11.0	10.9
1930	6.6	7.3	6.7	7.4	7.0
1940	6.5	8.6	7.1	7.8	7.5
1950	7.9	7.3	7.8	7.8	7.7
1960	6.9	5.6	6.6	6.7	6.4
1970	4.4	7.3	5.0	6.0	5.6
1980	8.4	7.9	8.3	8.5	8.2
			MALPE		
1920	5.2	14.1	7.1	7.9	8.6
1930	−1.4	1.2	−0.6	−0.8	−0.4
1940	3.5	5.9	4.4	3.8	4.4
1950	−6.8	−5.9	−6.6	−6.6	−6.5
1960	−5.2	−2.6	−4.5	−4.7	−4.2
1970	1.3	5.4	2.4	2.0	2.8
1980	−3.7	−1.7	−3.1	−3.3	−2.9

A constant mean is a necessary but not a sufficient condition to establish the stability of a distribution over time. Measures of the dispersion of errors around the mean must also be considered. Table 4 shows the standard deviations for absolute and algebraic forecast errors for each of the seven sets of 10-year projections. For all five techniques (especially Expo), standard deviations were large for 1920 but considerably smaller thereafter. There were no major differences among techniques regarding the size of standard deviations, nor was there any apparent trend over time. In all instances, standard deviations were larger for algebraic than absolute errors because algebraic errors account for both size and direction of error.

We are not primarily concerned in this study with differences in errors among alternative projection techniques. Consequently, the analysis throughout the remainder of the article will focus on projections coming solely from the Ave technique. An analysis of errors from the other four techniques was also performed (not shown here), and the results were similar to those reported here for Ave.

The stem-and-leaf plots (Tukey, 1977) in Figure 1 show the distributions of algebraic forecast errors for the seven sets of 10-year projections for the Ave technique. For each target year, the distribution is mound shaped and symmetric, except for occasional outliers (e.g., errors of 76 percent in 1920 and 40 percent in 1930). All seven distributions appear to be roughly normal. In fact, formal tests of normality (Shapiro and Wilk, 1965) reveal that the null hypothesis of a normal distribution cannot be rejected at a 5 percent level of significance for any target years except the two earliest, 1920 and 1930. Stem-and-leaf plots for absolute forecast errors were also generated. Not surprisingly, they revealed skewed (nonnormal) distributions, truncated at zero. (The interested reader can use the data shown in Fig. 1 to construct plots for absolute forecast errors by disregarding the signs of the errors and starting the scale at zero.)

To evaluate the stability of error distributions over time, we must focus on both means and variances. With θ representing the parameter of interest (mean or variance), we test three alternative hypotheses regarding stability:

Table 4. Standard Deviations for 10-Year Projections, by Technique and Target Year

Target year	Standard deviations				
	Line	Expo	Shift	Share	Ave
	Absolute Percentage Errors				
1920	10.7	23.4	11.9	14.8	15.1
1930	7.0	8.7	7.2	8.0	7.6
1940	5.0	7.5	5.3	5.9	5.8
1950	7.0	6.5	6.9	6.8	6.8
1960	6.3	5.0	6.0	5.5	5.5
1970	4.1	7.9	4.4	5.5	5.0
1980	5.5	5.8	5.5	5.8	5.5
	Algebraic Percentage Errors				
1920	12.6	24.4	13.5	16.7	16.6
1930	9.5	11.3	9.9	10.9	10.3
1940	7.5	9.8	7.8	9.1	8.5
1950	8.2	7.8	8.1	8.0	8.0
1960	7.8	7.1	7.7	7.2	7.3
1970	5.9	9.3	6.3	7.9	7.0
1980	9.4	9.7	9.5	9.9	9.5

1920 Stem	1920 Leaf	1930 Stem	1930 Leaf	1940 Stem	1940 Leaf	1950 Stem	1950 Leaf
4	9 (High = 76)	4		4		4	
4	11	4	0	4		4	
3	7	3		3		3	
3		3		3		3	
2	677	2		2		2	
2		2	4	2	013	2	
1	558	1		1	788	1	
1	1	1	003	1	01111123	1	
0	55555688999	0	55566667	0	566778899	0	579
0	111222223444	0	22223334	0	001223334444	0	1244
−0	0011222233	−0	000112233333444	−0	0111223	−0	01122222233333344
−0	6889	−0	55557777	−0	566778	−0	5666678889
−1	2	−1	02	−1		−1	0000112344
−1		−1	78	−1	34	−1	899
−2		−2	24	−2		−2	2
−2		−2		−2		−2	5
−3		−3		−3		−3	2

1960 Stem	1960 Leaf	1970 Stem	1970 Leaf	1980 Stem	1980 Leaf
4		4		4	
4		4		4	
3		3		3	
3		3		3	
2		2	5	2	
2		2		2	
1		1	7	1	55
1	034	1	00234	1	0124
0		0	556667899	0	57889
0	111122234	0	0111122222333333444	0	133344
−0	0122222333444	−0	011223344	−0	1122334444
−0	5555555667778888889	−0	679	−0	566777777899
−1	122	−1	0	−1	00001
−1	69	−1	5	−1	5577
−2	33	−2		−2	0
−2		−2		−2	9
−3		−3		−3	

Figure 1.—Stem-and-Leaf Plots for Algebraic Percentage Errors (Ave) for 10-Year Projections, by Target Year. The numbers for the stem indicate the first digit for each forecast error, and the numbers for the leaf indicate the second digit (e.g., an error of 10 percent is indicated by stem = 1 and leaf = 0). Each plot shows 50 errors, 1 for each state.

1. H_1: $\theta_{1920} = \theta_{1930} = \cdot \cdot \cdot = \theta_{1980}$,
2. H_2: $\theta_t = \theta_a$,
3. H_3: $\theta_t = \theta_{t-10}$,

where $t = 1920, 1930, \ldots, 1980$ and a = all target years between 1920 and 1980 (except t) combined.

H_1 tests the hypothesis that means (variances) of forecast errors are identical for all seven target years (1920–1980). This is the most stringent hypothesis. H_2 tests the hypothesis that the mean (variance) of forecast errors for target year t is identical to the mean (variance) for the distribution of errors from all six other target years combined. Seven different tests were conducted, one for each target year. H_3 tests the hypothesis that the mean (variance) of forecast errors for target year t is identical to the mean (variance) for the previous target year (t − 10). Six different tests were conducted, one for each pair of consecutive target years.

Because of the nonnormal distributions of algebraic errors for 1920 and 1930 and absolute errors for all target years, tests of means were conducted using well-known nonparametric alternatives to the traditional analysis of variance F test. The Kruskal–Wallis H test was used to test equality of means for H_1, and the two-sample Wilcoxon rank sum test was used for H_2 and H_3. For testing the equality of variances, Conover, Johnson, and Johnson (1981) showed that a variation of Levene's (1960) approximate F test is one of the most powerful and robust procedures. We chose this modified Levene test (in which the modification involved replacing the sample mean with the sample median) to test for the equality of variances.

The results of these tests are summarized in Table 5, which shows the p values associated with each of the three hypotheses. $p \geq 0.05$ indicates that the hypothesis of equal means (or variances) cannot be rejected at a 5 percent level of significance. Looking first at absolute forecast errors, the top panel of Table 5 shows that we cannot reject the hypothesis (H_1) that the mean errors for all seven sets of projections were identical, but we must reject the hypothesis that all seven variances were identical. On this most stringent test, then, the hypothesis regarding stability of absolute forecast errors must be rejected for one of the two parameters under consideration.

For H_2 and H_3 the results are more favorable for the notion of stability over time in the distribution of absolute forecast errors. The hypothesis (H_2) that the mean (variance) for target year t is identical to the mean (variance) for all other years combined must be rejected for only one out of seven sets of projections (1980 for the mean and 1920 for the variance). Even greater stability is found when testing H_3, that the mean (variance) for one target year is identical to the mean (variance) for the previous target year. This hypothesis must be rejected for only one out of six target years for means (1980) and cannot be rejected for any target years for variances.

With respect to algebraic forecast errors, the evidence regarding stability is much

Table 5. p Values for Tests of Stability of Means and Variances Over Time for 10-Year Ave Projections

Hypothesis	Absolute percentage error: Means	Absolute percentage error: Variances	Algebraic percentage error: Means	Algebraic percentage error: Variances
H_1	0.13	<0.01	<0.01	0.02
H_2				
1920	0.92	<0.01	<0.01	0.01
1930	0.42	0.56	0.81	0.55
1940	0.46	0.66	<0.01	0.51
1950	0.63	0.95	<0.01	0.20
1960	0.42	0.21	<0.01	0.03
1970	0.05	0.12	<0.01	0.03
1980	0.02	0.33	0.02	0.93
H_3				
1930	0.63	0.05	<0.01	0.16
1940	0.29	0.89	<0.01	0.83
1950	0.90	0.71	<0.01	0.53
1960	0.35	0.25	0.25	0.41
1970	0.41	0.71	<0.01	0.94
1980	<0.01	0.55	<0.01	0.03

Note: H_1: Means (variances) are identical for all seven target years. H_2: Mean (variance) for target year t is identical to mean (variance) from distribution for all other target years combined. H_3: Mean (variance) for target year t is identical to mean (variance) for target year $t - 10$. $p \geq 0.05$ indicates that the hypothesis cannot be rejected at a 5 percent level of significance.

weaker. For H_1 we must reject the hypotheses that both the means and variances from all seven sets of projections were identical. For H_2 the hypothesis of equality of means must be rejected for six of the seven comparisons and the hypothesis of equality of variances must be rejected for three of the seven comparisons. For H_3 the hypothesis of equality of variances must be rejected for only one of the six target years (1980), but the hypothesis of equality of means must be rejected for five of the six target years (all except 1960).

We replicated these tests for states divided into size categories (<1 million, ≧1 million) and growth-rate categories (<20 percent, ≧20 percent).[2] The results were much the same. Although means and variances were considerably different for small states than for large states and for rapidly growing states than for slowly growing states, within each size or growth-rate category, the hypothesis of equality of means and variances for absolute errors could not be rejected in most instances. For algebraic errors, however, the hypothesis of equality of variances had to be rejected in a number of instances and the hypothesis of equality of means had to be rejected in almost all instances.

We also replicated the tests for the 20-year projections with a 20-year base period. (For projections with a 20-year horizon, a 20-year base period provided somewhat greater accuracy and stability than a 10-year base period.) Although the means and standard deviations for 20-year projections were approximately twice as large as for 10-year projections, the results regarding stability over time for absolute and algebraic errors were virtually the same as those reported in Table 5.

We conclude from these results that there was a high degree of stability over time for both the means and variances of absolute forecast errors for state projections during the 20th century. There was somewhat less stability for the variances of algebraic forecast errors and no stability at all for mean algebraic forecast errors. Since the MALPE is a measure of bias, this tells us that the study of past forecast errors cannot help us predict the overall tendency for a current set of projections to be too high or too low.[3] A number of other studies have drawn similar conclusions (e.g., Kale et al., 1981; Smith, 1984; Stoto, 1983). The stability for both the means and variances of absolute forecast errors, however, provides evidence that the study of past forecast errors may help us predict the level of accuracy of current population projections, even if we cannot predict their bias.

Empirical Confidence Limits

The second objective of this study is to evaluate the validity of using data on the distribution of past forecast errors to create confidence intervals for population projections. The results presented in the preceding section suggest that the critical assumption underlying the construction and use of empirical confidence intervals—that the distribution of errors remains stable over time—is reasonably well satisfied in this sample for absolute forecast errors. How can these results be used to construct and test confidence intervals?

Williams and Goodman (1971) suggested a method for constructing "empirical confidence limits" based on the distribution of past forecast errors. We have modified and adopted this technique because it can accommodate any error distribution, including the asymmetric and truncated distributions of absolute forecast errors found in this sample. The Williams and Goodman approach also permits an assessment of the confidence limits; that is, we can compare the actual number of errors falling inside the limits with the expected number.

For each of the seven sets of 10-year projections, absolute forecast errors were ranked and the 90th percentile error (i.e., the absolute percentage error that was larger than exactly 90 percent of all absolute percentage errors) was determined. The 90th percentile errors for each target year are shown in the second column of Table 6. We followed three different approaches in using these errors to construct 90 percent confidence limits:

Table 6. Actual and Predicted 90th Percentile Errors for Ave Technique, by Target Year

		Distribution of errors from previous target year		Distribution of errors from all other target years		Distribution of errors from all other target years (excl. 1920)	
Target year	Actual 90th percentile error	Predicted 90th percentile error	% of actual errors less than predicted error	Predicted 90th percentile error	% of actual errors less than predicted error	Predicted 90th percentile error	% of actual errors less than predicted error
1920	36.4	—	—	15.1	82	—	—
1930	17.8	36.4	98	16.6	88	15.0	88
1940	17.6	17.8	92	16.7	90	15.0	88
1950	18.8	17.6	88	16.6	88	15.0	88
1960	14.3	18.8	94	17.4	94	15.3	92
1970	13.2	14.3	92	17.7	98	16.6	96
1980	15.1	13.2	82	17.4	94	15.2	92

1. The 90th percentile error from target year $t - 10$ was used as a forecast of the 90th percentile error in target year t.

2. The 90th percentile error from the distribution of errors from all other target years was used as a forecast of the 90th percentile error in target year t.

3. The 90th percentile error from the distribution of errors from all other target years (excluding 1920) was used as a forecast of the 90th percentile error in target year t.

The first two approaches were selected to correspond to null hypotheses H_3 and H_2, as described in the preceding section. The third approach was selected to eliminate the effects of several unusually large outliers in the error distribution for 1920. If the distribution of absolute forecast errors remains stable over time, 90th percentile errors based on past error distributions (or in the case of the second and third approaches, error distributions both preceding and following the target year) should provide accurate predictions of the 90th percentile error for the target year. Since we expect 90 percent of the absolute forecast errors to fall below this value, the predicted 90th percentile error for any target year may be thought of as a one-sided (upper) confidence limit with an associated confidence coefficient of 0.90.

To assess the accuracy of these three approaches to the construction of empirical confidence limits, we compared the predicted with the actual 90th percentile errors for each target year from 1920 to 1980 and computed the proportion of absolute forecast errors that fell within the predicted value. These results are shown in the last six columns of Table 6. In general, the results are quite satisfactory, especially in light of the small sample size. In most instances the proportion of errors falling within the predicted confidence limit was close to 90 percent. Using data from all other target years generally provided a more accurate confidence limit than using data solely from the previous target year, and excluding the first set of projections (target year 1920) further improved accuracy. For this third approach, between 88 percent and 92 percent of errors fell within the predicted 90 percent value for all target years except 1970, when 96 percent were within the predicted value.[4]

These empirical confidence limits appear to be neither too wide nor too narrow. For some target years more than 90 percent of the errors fell within the predicted value, and for other years fewer than 90 percent fell within the predicted value. The target years with too many errors within the predicted values essentially offset the years with too few. Summing over all target years, 91.0 percent of actual errors fell within the predicted value using the first approach to confidence limits, 90.6 percent using the second approach, and 90.7 percent using the third approach.

The same three approaches to empirical confidence limits were applied to 20-year projections with a 20-year base period (not shown). Although the 90th percentile errors were larger than for the 10-year projections, the relationships between actual and predicted errors were similar to those shown in Table 6.

We also constructed empirical confidence limits for states grouped by population size (<1 million, ≧1 million) and growth rate (<20 percent, ≧20 percent). The results for large states (≧1 million) and slowly growing states (<20 percent) were similar to the results shown in Table 6: Actual 90th percentile errors corresponded very closely to the predicted ones. For small states (<1 million) and rapidly growing states (≧20 percent), however, predicted 90th percentile errors differed considerably from actual 90th percentile errors. These less satisfactory results can be attributed to the more volatile nature of population growth in small and rapidly growing states and—perhaps more important—to the very small sample sizes for these categories (between 5 and 19 states per decade).

Conclusions

Can confidence intervals really be made for population projections? Under the formal definition of confidence intervals, the answer is no: Confidence intervals cannot be constructed because the probability distribution of future forecast errors is unknown (and unknowable) at the present time. It is possible, however, to look at forecasts made in the past to see how accurate they were in predicting population change. If current projection techniques are similar to those used in the past, and if the degree of uncertainty is about the same in the future as it was in the past, we can assume that future forecast errors will be drawn from the same distribution as past errors (Keyfitz, 1981:587). Data on the distribution of past forecast errors can therefore be used to construct empirical confidence limits for population projections.

The critical assumption underlying the construction and use of empirical confidence limits is that the distribution of forecast errors remains stable over time. The present study has shown that the distribution of absolute forecast errors for states remained relatively stable over the decades of the 20th century (especially since 1920). This finding is powerful, given the tremendous fluctuations in population growth caused by world wars, depressions, baby booms, and countless other events during this century. It gives us reason to believe that just as data on past population trends can frequently provide reasonably reliable short- to medium-term population forecasts, so might data on past forecast errors be able to provide useful predictions of the distribution of future forecast errors. Indeed, the empirical confidence limits tested in this study were generally successful in predicting the range of future errors.

This study has also shown that the distribution of algebraic forecast errors was not at all stable over the course of the 20th century. Although standard deviations remained moderately stable over time, mean algebraic errors varied dramatically and unpredictably from one decade to the next. In this sample, then, data on the bias of past projections did not provide a reliable basis for predicting the bias of current projections. The approach taken in this study is not likely to be useful for predicting the overall tendency for a particular set of population projections to be too high or too low.

There are many different ways to produce confidence intervals for population projections (e.g., Cohen, 1986; Land, 1986). Each approach has unique characteristics, and within each approach, different projection techniques imply different confidence intervals as well. Future research must therefore evaluate various approaches for constructing confidence intervals, focusing on the effects of using different projection techniques, base periods, projection horizons, and so forth. The results of this study imply that such research will be very useful. Even though we may never be able to forecast future *populations* with a high degree of accuracy, we may be able to develop relatively accurate forecasts of the *distribution of errors* surrounding our point forecasts. Indicating the potential range of errors surrounding population forecasts may be the most useful service the producers of population projections can provide to their users.

Notes

[1] The large errors for the 1920 projections were caused primarily by the large numbers of small states and states with high growth rates. There were 13 states with populations of less than 500,000 in 1910 and 11 states that grew by more than 50 percent between 1900 and 1910.

[2] It would also be interesting to replicate these tests for states divided by size of place and rate of growth simultaneously. Smith (1987) reported some important interactions between these two variables. Unfortunately, the sample size used in this study is too small to permit such an analysis.

[3] This refers to bias for an entire set of projections. There is evidence that extrapolative projection techniques have predictable biases for places with extreme growth rates during the base period (Smith, 1987).

[4] We also looked at 50th, 75th, and 95th percentile errors. For 95th percentile errors, the results were similar to those reported in Table 6. For 50th and 75th percentile errors, however, there were larger differences between the actual and predicted numbers of errors falling below the predicted value than shown in Table 6. We believe this is because errors are more tightly bunched around the 50th and 75th percentile errors than around the 90th and 95th percentile errors. Therefore relatively small differences between the actual and predicted values of percentile errors can cause relatively large differences in the number of errors falling below the predicted value.

References

Ascher, W. 1978. *Forecasting: An Appraisal for Policy-Makers and Planners.* Baltimore, Md.: Johns Hopkins University Press.

Cohen, J. 1986. Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography* 23:105–126.

Conover, W. J., M. E. Johnson, and M. M. Johnson. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23:351–361.

Dorn, H. F. 1950. Pitfalls in population forecasts and projections. *Journal of the American Statistical Association* 45:311–334.

Federal–State Cooperative Program for Population Projections. 1984. *State Survey on Population Projections.* Sacramento: California Dept. of Finance, Population Research Unit.

Greenberg, M. 1972. A test of combination of models for projecting the population of minor civil divisions. *Economic Geography* 48:179–188.

Irwin, R. 1977. *Guide for Local Area Population Projections*, Technical Paper No. 39. Washington, D.C.: U.S. Bureau of the Census.

Isserman, A. 1977. The accuracy of population projections for subcounty areas. *Journal of the American Institute of Planners* 43:247–259.

Isserman, A., and P. Fisher. 1984. Population forecasting and local economic planning: The limits on community control over uncertainty. *Population Research and Policy Review* 3(4):27–50.

Kale, B., P. Voss, C. Palit, and H. Krebs. 1981. On the Question of Errors in Population Projections. Paper presented at the Annual Meeting of the Population Association of America, Washington, D.C.

Keyfitz, N. 1972. On future population. *Journal of the American Statistical Association* 67:347–363.

———. 1981. The limits of population forecasting. *Population and Development Review* 7:579–593.

Land, K. C. 1986. Methods for national population forecasts: A review. *Journal of the American Statistical Association* 81:888–901.

Lee, R. D. 1974. Forecasting births in post-transition populations: Stochastic renewal with serially correlated fertility. *Journal of the American Statistical Association* 69:607–617.

Levene, H. 1960. Robust tests for equality of variances. Pp. 278–292 in I. Olkin (ed.), *Contributions to Probability and Statistics.* Palo Alto, Calif.: Stanford University Press.

Murdock, S. H., F. L. Leistritz, R. R. Hamm, S.-S. Hwang, and B. Parpia. 1984. An assessment of the accuracy of a regional economic-demographic projection model. *Demography* 21:383–404.

Saboia, J. 1974. Modeling and forecasting population by time series: The Swedish case. *Demography* 11:483–492.

Schmitt, R., and A. Crosetti. 1951. Accuracy of the ratio method for forecasting city population. *Land Economics* 27:346–348.

———. 1953. Short-cut methods of forecasting city population. *Journal of Marketing* 17:417–424.

Shapiro, S. S., and M. B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591–611.

Siegel, J. S. 1953. Forecasting the population of small areas. *Land Economics* 29:72–88.

———. 1972. Development and accuracy of projections of population and households in the United States. *Demography* 9:51–68.

Smith, S. K. 1984. *Population Projections: What Do We Really Know?*, Monograph No. 1. Gainesville, Fla.: University of Florida, Bureau of Economic and Business Research.

———. 1987. Tests of forecast accuracy and bias for county population projections. *Journal of the American Statistical Association* 82:991–1003.

Stoto, M. 1983. The accuracy of population projections. *Journal of the American Statistical Association* 78:13–20.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, Mass.: Addison-Wesley.

U.S. Bureau of the Census. 1957. *Illustrative Projections of the Population, by State, 1960, 1965 and 1970*, Current Population Reports, Ser. P-25, No. 160. Washington, D.C.: U.S. Government Printing Office.

———. 1983. *Number of Inhabitants, 1980 Census of Population* (PC80-1-A1). Washington, D.C.: U.S. Government Printing Office.

U.S. Bureau of Economic Analysis. 1974. *Area Economic Projections 1990*. Washington, D.C.: U.S. Government Printing Office.

Voss, P., C. Palit, B. Kale, and H. Krebs. 1981. *Forecasting State Population Using ARIMA Time Series Techniques*. Madison: University of Wisconsin, Applied Population Laboratory.

White, H. R. 1954. Empirical study of the accuracy of selected methods of projecting state population. *Journal of the American Statistical Association* 49:480–498.

Williams, W. H., and M. L. Goodman. 1971. A simple method for the construction of empirical confidence limits for economic forecasts. *Journal of the American Statistical Association* 66:752–754.