

Evaluating Forecast Error in State Population Projections Using Census 2000 Counts

By Paul R. Campbell

Population Division
U.S. Bureau of the Census
Washington, D.C. 20233

July 2002

Population Division Working Paper Series No. 57

Evaluating Forecast Error in State Population Projections Using Census 2000 Counts

Paul R. Campbell

Population Division
U.S. Bureau of the Census
Washington, D.C. 20233

Abstract

In order to determine if a popular summary statistic, the mean absolute percentage error (MAPE), is a valid measure of forecast error for the Census Bureau's 1995 to 2000 state population projections, statistical tests and graphs were used to determine if the error distribution is strongly influenced by outliers. It was found that the absolute percentage error distribution is skewed and asymmetrical. Since the MAPE understates accuracy, MAPE-R, a variant of MAPE derived from the transformed absolute percentage error distribution was accepted as more accurate. Using simple extrapolated projections as a standard to compare forecast error, the findings suggest that the Census Bureau's projections are fairly accurate over a short projection horizon.

Paper Presented at the Federal Forecasters Conference in Washington, D.C. April 18, 2002.

Disclaimer: This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

EVALUATING FORECAST ERROR IN STATE POPULATION PROJECTIONS USING CENSUS 2000 COUNTS

Paul R. Campbell, U.S. Bureau of the Census

Introduction. Users of the Census Bureau's 1990 census-based state population projections for 1995 to 2025 are interested in the Bureau's forecast error. Even counties or smaller geographic areas are dependent upon state level accuracy, since projections for these areas often are prorated to the state figures. In this study, Census 2000 counts are used to measure forecast error in projections for April 1, 2000. This is the first opportunity to evaluate the 1995 to 2025 projections with the 'truth' assuming that the Census 2000 results are correct. This evaluation specifically examines the forecast error of the 2000 state population projection totals (including the District of Columbia) for Series A and B, and a simple extrapolated projection to identify the more accurate set of projections and the state outliers. The basic statistical method used to detect state population projection error is to examine the Percentage Error (PE) and the Absolute Percentage Error (APE), while the overall accuracy of the set of projections is measured by the Mean Absolute Percentage Error (MAPE). Additionally, this study examines the skewness and asymmetry of the APE distribution for states to determine if there is a need to use MAPE-R, derived from the transformed APEs, as recommended in the literature in order to correct for the influence of outliers on the mean.

Several questions are addressed in this state projections evaluation. How accurate are the Census Bureau's state population projections for the year 2000? Which Census Bureau state population projection series is the most accurate? Are the Census Bureau's projections as good as or better than the results obtained from simple extrapolated projections? To answer these questions, a discussion on evaluating the Census Bureau's state population projections is presented in several sections. First, the "Prior Research" section reviews recent literature on the evaluation of subnational population projections. The "State Projections for 2000" section discusses the methodology of the Census Bureau's state population projections for 1995 to 2025. The "Extrapolated Projections" section identifies the procedures used to produce a simple extrapolated projection from the enumerated 1990 census counts and the post-1990 census estimates for 1995. Next, the "Calculation of MAPE" section explains how the basic statistical summary measures are derived. The "Transformed APEs and Test for Symmetry" section (1) describes how the APE distribution is examined to see if it is asymmetrical and needs correcting, and (2)

presents the necessary transformation and conversion formula used to correct for the influence of extreme outliers. Finally, the results are discussed in the "Findings and Conclusions" section.

Prior Research. Since literature on the evaluation of state-level population forecasts is not extensive, a broader review of research on the evaluation of subnational population estimates and projections provide useful guidelines for measuring forecast errors. Most frequently subnational population estimates or projections were evaluated using the mean absolute percentage error (MAPE), as indicated in evaluations by Swanson, Tayman, and Barr (2000), Tayman and Swanson (1999), Campbell (1997), Davis (1994), Smith and Sincich (1992). The MAPE is a measure of the central tendency of errors calculated by averaging the sum of the percentage differences between the projections and the census for states, ignoring the plus or minus sign.

There are other alternative statistical measures used to evaluate errors (or differences) in subnational population estimates and projections. Davis (1994), evaluating post-1980 county populations estimates with the 1990 census used several summary measures. In addition to the MAPE, he used the mean algebraic percentage error (MALPE), weighted mean absolute percentage error (WMAPE), root mean square error (RMSE), index of dissimilarity (INDISS), median percent difference, and 90th percentile (or percent error at which 90 percent of the observations are lower). A complete description of these and other statistical measures is available in his study and in Armstrong (1977). Davis (1994) limits his discussion of findings to the "familiar" MAPE, since this summary statistic is highly correlated with the other summary measures for most of his tabulations, except for the MALPE and the 90th percentile.

In questioning the validity of the MAPE measure, Tayman and Swanson (1999) use the MAPE, the Symmetrical MAPE (SMAPE), and a class of measures known as Minimization-Estimators (M-estimators) to evaluate county forecasts for selected states. The M-estimators are described by Tayman and Swanson (1998:303) as "minimizing a more general objective function using maximum likelihood procedures rather than the sum of squared residuals associated with the sample mean, a sum that is highly sensitive to outliers."

Their findings suggest that a robust M-estimator like the Tukey-M statistic is a suitable alternative summary measure of forecast error.

Swanson, et al., (2000) in evaluating subnational estimates, argues that the MAPE is reliable, easy to interpret, and clear in its presentation. On the other hand, they acknowledge that the MAPE in many cases lacks validity, since the APEs used in its calculation are right-skewed, such that extreme values can unduly influence the MAPE. To reduce the effects of outliers and asymmetrical distributions on the arithmetic mean, the alternatives are the median, the geometric mean, the weighted mean, the M-estimator, and data transformation.

In their evaluation of county estimates, Swanson, et al., (2000) and Tayman and Swanson (1999) suggest that some extreme outliers may influence the MAPE by pulling its value upward so that it is not valid for a data set. They suggest validating the MAPE with tests for skewness and symmetry. If the MAPE is not valid, then a variant of the MAPE is calculated using a data transformation that corrects for the error inflation due to the outliers.

This paper follows guidelines found in evaluation literature that recommend performing a data transformation to obtain a variant of the MAPE whenever the original distribution of APE is not symmetrical. In order to test for the effect that outliers have on the summary measure, Swanson, et al., used a modified Box-Cox (1964) transformation to obtain a symmetrical distribution of the original APEs, such that very large errors (or outliers) are compressed. The original distribution can be tested for asymmetry using graphic devices like histograms and boxplots; or statistical measures like the skewness coefficients (Snedecor and Cochran 1980:78); and the D'Agostino skewness test (D'Agostino, et al., 1990). When the original APEs are biased upward, then Swanson, et al., (2000) recommend calculating a summary measure from the transformed APEs that they refer to as MAPE-T. Since the MAPE-T, the average of the transformed APEs, is not represented in a familiar scale, they recommend using a nonlinear power function to statistically map the scales of the original error observations to the transformed observations. Swanson, et al., (2000) suggest calculating a re-expressed average (MAPE-R) in metrics, which is solved using linear regression results and the logarithm of MAPE-T. They argue that MAPE-R is a measure of central tendency of the error that is not influenced by the asymmetry and outliers that characterize the untransformed absolute percentage error distribution. Evaluating county

estimates, Swanson, et.al. (2000:199) validated the MAPE-R statistic by finding consistent results using Tukey-M.

Swanson, et al., (2000) identified some of the shortcomings of the alternative summary measures that may more accurately describe APEs as follows: M-estimators are not easy to explain, the median is "influenced by changes in the centermost observations resulting from grouping," the geometric mean is affected by the logarithmic transformation which sometimes fails to yield a distribution that has optimal symmetry, and the loss function lacks a standard weighting scheme.¹ Additionally, they acknowledge that MAPE-R was cumbersome to calculate and required the use of different statistical software packages.

From a different perspective, economists Kolb and Stekler (1993) recommend using mean square error (MSE) to test whether a set of economic forecasts are statistically significant or better than "naïve" or "no change" forecasts. Using the simple extrapolated population difference as a standard, they used Theil's U statistic to determine if the more complex projection model performs at least as well as the simplest model. Furthermore, evaluating state population projection models that ranged from simple to complex, Smith and Sincich (1992) concluded that the simple trend projections derived from linear extrapolation are just as accurate as the more complex models like the Census Bureau's cohort-component projections. This analysis goes a step further than Smith and Sincich (1992) or Campbell (1997), the MAPE or MAPE-R from the Census Bureau's 2000 state projection and the census population is compared to the same summary statistics from a standard, the simple extrapolated population projection and the census population.

State Projections for 2000. The Census Bureau's state population projections use detailed demographic accounting procedures and professional judgement in developing projection assumptions. The Census Bureau's state projections were prepared for July 1 of each year from 1995 to 2025 using the cohort-component projection method. The cohort-component method is based on the traditional demographic accounting system:

$$P_1 = P_0 + B - D + DIM - DOM + IM - EM$$

where: P_1 = population at the end of the period, P_0 = population at the beginning of the period, and the

¹For a discussion on the merits of the loss function versus MAPE / MAPE-R and other summary measures see Coleman (2000 and 2002).

following events during the period: B = births, D = deaths, DIM = domestic in-migration, DOM = domestic out-migration, (both DIM and DOM are aggregations of the state-to-state migration flows), IM = immigration, and EM = emigration.

Each component of population change -- births, deaths, internal migration (domestic or state-to-state migration flows), and international migration (immigration and emigration) -- utilizes separate projection assumptions for each birth cohort by single year of age, sex, race, and Hispanic origin. The race and Hispanic origin groups projected separately were: non-Hispanic White; non-Hispanic Black; non-Hispanic American Indian, Eskimo, and Aleut; non-Hispanic Asian and Pacific Islander; Hispanic White, Hispanic Black, Hispanic American Indian, Eskimo, and Aleut; Hispanic Asian and Pacific Islander. The detailed components used in the state population projections were derived from vital statistics, administrative records, 1990 census data, state population estimates (U.S. Bureau of the Census, 1996c), and the middle series of the national population projections (Day, 1996). Detailed assumptions and procedures by which these data were generated by single year of age, sex, race, and Hispanic origin are described in detail in the report, "Population Projections for States, by Age, Sex, Race, and Hispanic Origin: 1995 to 2025," (Campbell, 1996). Overall, the assumptions concerning the future levels of fertility, mortality, and international migration are consistent with the assumptions developed for the national population projections (Day, 1996).

Once separate data components were developed, the cohort-component method was applied, producing the detailed demographic projections. For the start of each projection year, the beginning population for each state was disaggregated into race and Hispanic origin categories (the eight groups previously identified) by sex and single year of age (0 to 84, and 85 plus). Components of change were individually applied to each group to project the next year's population. For the mortality component, each age-sex-race/ethnic group was survived forward one year using the pertinent survival rate. The internal redistribution of the population was accomplished by applying the appropriate state-to-state migration rates to the survived population in each state. The projected out-migrations were subtracted from the state of origin and added to the state of destination (as in-migrants). Next, the number of immigrants from abroad was added to each state, while the number of emigrants leaving each state was subtracted. Applying the appropriate age-race/ethnic-specific birth rates to females of childbearing age created the populations less than one

year of age. The number of births by sex and race/ethnicity were survived forward and exposed to the appropriate migration rate to yield the population less than one year of age. The results were adjusted to be consistent with the national population projections by single years of age, sex, and race/ethnicity. Both the state and national population projection reports indicate that 1994 was the base year or the most recent year estimates were used to begin the forecast. However, the first year of the states projection horizon, 1995, was also adjusted to be consistent with a set of preliminary 1995 state estimates only available by age and sex. The entire process was then repeated for each year of the projection.

Two sets of state population projections were prepared and the only component specified differently in each projection model was the domestic migration component. The dynamic possibilities of change in state-to-state migration make it the most difficult component to forecast. Migration trends in the Census Bureau's state projections are based on matched Internal Revenue Service (IRS) individual income tax return data sets containing 19 annual observations (from 1975-76 to 1993-94) on each of 2,550 state-to-state migration flows. The two projection series provide users with different domestic migration scenarios since one set includes the rate of change in employment. Both sets of state projections were summed and adjusted by age, sex, race and Hispanic origin to agree with the middle series of the national population projection.

The Census Bureau refers to Series A, which uses a time series model, as the "preferred series." The first five years of the projection horizon (1995 to 2000) use the time series projection exclusively. The next ten years on the projection horizon (2000 to 2010) are interpolated toward the mean of the series, while the final 15 years (2010 to 2025) use the mean of the series exclusively. Series B is the economics model. State-to-state migration flows are derived from the Bureau of Economic Analysis projected rate of change in employment in the origin and the destination states. The "preferred series" was accepted as the projection model most likely to be the more accurate series based on results from ex-post facto evaluations.²

The current sets of state population projections were previously evaluated in the Census Bureau using post-

² Ideally, the "preferred series" should be based on the evaluation of preliminary projections where the most recent launch year estimates are withheld from the projection so that estimates can be compared against the preliminary projections and prior evaluation results.

1990 census estimates for July 1, 1996. The MAPEs calculated for 1996 were found to be fairly accurate (at or below the U.S. total of 0.40 percent for Series A and 0.30 percent for Series B, for all regions except the West). That earlier evaluation also looked separately at the components of change and found that the birth component was the most accurate followed by the mortality component. The study concluded that both domestic and international migration components were more difficult to forecast accurately, and domestic migration was the least accurate component in the projections (Campbell, 1997).

An important first step in this evaluation was to obtain projections that are consistent with the Census 2000 reference date. While projections are available annually for July 1 of each year, there are none readily available for the target date centered on April 1, 2000. To match the projections to the census date, the solution was to linearly interpolate between the July 1, 1999 and July 1, 2000 state population projection totals to obtain projections for the census date April 1, 2000 using Waring's formula (see Shryock and Siegel, 1976:533).³

Extrapolated Projections. A simple extrapolated total population projection for each state was used as a standard to evaluate the forecast error in the state's total population projections. The extrapolated state population projections were derived by linearly extrapolating from the enumerated April 1, 1990 census population and the July 1, 1995 populations⁴ to April 1, 2000 for every state. Smith and Sincich (1992), using several techniques in an evaluation which ranged from extrapolating growth rates and ratio shares to time series models, concluded that the linear extrapolation and ratio share models performed the best. Based on their recommendation, the following formula was used to extrapolate to April 1, 2000:

$$P_t = P_0 + X / Y(P_b - P_0)$$

where the P_t is the state population projection for the target year (April 1, 2000), P_0 is the state population size on July 1, 1995, P_b is the state population size in the base year (April 1, 1990), X is the number of years

in the base period (5.25 years between April 1, 1990 to July 1, 1995) and Y is the number of years in the projection horizon (4.75 years between July 1, 1995 to April 1, 2000). No attempt was made to control the extrapolated state totals to independently derived national projection totals, which results in simple extrapolations for states not affected by inflation/deflation errors (from states being forced to sum prorata to the nation).⁵

Calculation of MAPE. Forecasters tend to treat the terms projection, extrapolation, prediction, and forecast as synonymous (Armstrong, 2001:39). In this study, the initial forecast error refers to the percentage difference or error between a state's total population projection and the Census 2000 population enumerated for the same date.⁶ Calculating the percentage error (PE) is useful in examining the error magnitude, direction of error, and identifying outliers in the evaluation of the state projections with census counts (see Table 1). The absolute percentage error (APE) is calculated without regard to the direction of error. The statistical measure that summarizes the APE distribution is the mean absolute percentage error (MAPE). The formula for the APEs and MAPE, ignoring the $\pm$ sign, are as follows:

$$APE_i = |P_i - C_i| \div C_i * 100$$

$$MAPE = \Sigma(APE_i) \div N$$

where N refers to the number of states (in the U.S., a region, or a division), P is the projected or extrapolated population, C is the census population, and i refers to the state.

MAPEs were developed for the United States (the states and the District of Columbia), where N equals 51; and for each census region or division, where N equals the number of states in each region or division. All data evaluated are from unrounded state population projection figures reported in U.S. Bureau of the Census (2000, 1996a, 1996b, and 1996c) and Campbell (1996).

Transformed APEs and Test for Symmetry. After APE and MAPE are calculated, Swanson, et al., (2000:194) suggest using transformed APEs that are

³ Waring's two-point interpolation formula was used for each state: $f(x) = [f(a)*(x-b)/(a-b)] + [f(b)*(x-a)/(b-a)]$, where $f(x)$ = April 1, 2000 population; $f(a)$ = July 1, 1999 population; $f(b)$ = July 1, 2000 population; and the proportions of the year were $x = 2000+92/366$; $a = 1999+182/365$; and $b = 2000+183/366$. Wang (2002) using geometric interpolation to calculate the April 1, 2000 projections reported slightly different results.

⁴ The July 1, 1995 totals rather than July 1, 1994 totals were used since the state projections for the first target year 1995 were inflated/deflated to the preliminary 1995 state estimates.

⁵ Comparison of the April 1, 2000 U.S. extrapolated state population totals with the Census 2000 total indicated that the extrapolated projections underprojected the U.S. total population by 2.12 percent. This is more accurate than the Series A and B projections, which were underprojected by 2.62 percent.

⁶ Projection evaluations based on estimates are ephemeral, since estimates may be corrected several times during the intercensal decade, only to be finalized after incorporating the latest census results.

Table 1. Percentage Error in Census Bureau's Projections Series A and B, and Extrapolated Projections, 2000

Region, Division and State	Projection Series A			Projection Series B		Extrapolated Projections	
	Census	Number	Percentage Error	Number	Percentage Error	Number	Percentage Error
NORTHEAST							
<i>New England</i>							
Maine.....	1,274,923.....	1,258,277	-1.31	1,249,996	-1.96	1,253,593	-1.67
New Hampshire.....	1,235,786.....	1,220,918	-1.20	1,213,932	-1.77	1,183,513	-4.23
Vermont.....	608,827.....	615,398	1.08	606,141	-0.44	604,642	-0.69
Massachusetts.....	6,349,097.....	6,192,883	-2.46	6,216,274	-2.09	6,125,219	-3.53
Rhode Island.....	1,048,319.....	997,150	-4.88	988,592	-5.70	977,458	-6.76
Connecticut.....	3,405,565.....	3,283,684	-3.58	3,285,888	-3.51	3,263,436	-4.17
<i>Middle Atlantic</i>							
New York.....	18,976,457.....	18,144,490	-4.38	18,171,587	-4.24	18,267,830	-3.73
New Jersey.....	8,414,350.....	8,167,064	-2.94	8,174,604	-2.85	8,139,921	-3.26
Pennsylvania.....	12,281,054.....	12,196,852	-0.69	12,213,689	-0.55	12,243,940	-0.30
MIDWEST							
<i>East North Central</i>							
Ohio.....	11,353,140.....	11,311,801	-0.36	11,343,222	-0.09	11,425,043	0.63
Indiana.....	6,080,485.....	6,033,717	-0.77	6,047,746	-0.54	6,038,106	-0.70
Illinois.....	12,419,293.....	12,040,250	-3.05	12,057,487	-2.91	12,191,274	-1.84
Michigan.....	9,938,444.....	9,673,709	-2.66	9,703,959	-2.36	9,779,185	-1.60
Wisconsin.....	5,363,675.....	5,317,128	-0.87	5,314,560	-0.92	5,331,994	-0.59
<i>West North Central</i>							
Minnesota.....	4,919,479.....	4,819,724	-2.03	4,812,141	-2.18	4,821,686	-1.99
Iowa.....	2,926,324.....	2,897,210	-0.99	2,888,920	-1.28	2,900,595	-0.88
Missouri.....	5,595,211.....	5,530,291	-1.16	5,535,993	-1.06	5,510,330	-1.52
North Dakota.....	642,200.....	660,765	2.89	656,275	2.19	643,632	0.22
South Dakota.....	754,844.....	774,908	2.66	768,379	1.79	758,890	0.54
Nebraska.....	1,711,263.....	1,702,280	-0.52	1,697,153	-0.82	1,690,255	-1.23
Kansas.....	2,688,418.....	2,663,372	-0.93	2,669,488	-0.70	2,644,715	-1.63
SOUTH							
<i>South Atlantic</i>							
Delaware.....	783,600.....	765,340	-2.33	756,486	-3.46	763,341	-2.59
Maryland.....	5,296,486.....	5,264,225	-0.61	5,251,268	-0.85	5,278,573	-0.34
District of Columbia.....	572,059.....	524,104	-8.38	530,118	-7.33	506,624	-11.44
Virginia.....	7,078,515.....	6,979,511	-1.40	6,949,653	-1.82	7,008,299	-0.99
West Virginia.....	1,808,344.....	1,840,407	1.77	1,832,857	1.36	1,859,488	2.83
North Carolina.....	8,049,313.....	7,750,347	-3.71	7,761,108	-3.58	7,707,711	-4.24
South Carolina.....	4,012,012.....	3,849,116	-4.06	3,843,811	-4.19	3,842,137	-4.23
Georgia.....	8,186,453.....	7,843,520	-4.19	7,859,752	-3.99	7,854,730	-4.05
Florida.....	15,982,378.....	15,181,072	-5.01	15,197,736	-4.91	15,276,275	-4.42
<i>East South Central</i>							
Kentucky.....	4,041,769.....	3,988,406	-1.32	3,983,394	-1.44	4,018,483	-0.58
Tennessee.....	5,689,283.....	5,638,679	-0.89	5,648,707	-0.71	5,598,854	-1.59
Alabama.....	4,447,100.....	4,441,087	-0.14	4,426,738	-0.46	4,445,124	-0.04
Mississippi.....	2,844,658.....	2,810,204	-1.21	2,819,891	-0.87	2,809,485	-1.24
<i>West South Central</i>							
Arkansas.....	2,673,400.....	2,624,492	-1.83	2,615,811	-2.15	2,604,110	-2.59
Louisiana.....	4,468,976.....	4,420,298	-1.09	4,439,675	-0.66	4,453,055	-0.36
Oklahoma.....	3,450,654.....	3,367,555	-2.41	3,365,485	-2.47	3,397,212	-1.55
Texas.....	20,851,820.....	20,050,906	-3.84	20,106,138	-3.58	20,296,038	-2.67
WEST							
<i>Mountain</i>							
Montana.....	902,195.....	946,058	4.86	933,518	3.47	934,667	3.60
Idaho.....	1,293,953.....	1,338,081	3.41	1,324,245	2.34	1,304,869	0.84
Wyoming.....	493,782.....	522,521	5.82	517,023	4.71	504,236	2.12
Colorado.....	4,301,261.....	4,149,286	-3.53	4,135,073	-3.86	4,155,703	-3.38
New Mexico.....	1,819,046.....	1,852,019	1.81	1,849,867	1.69	1,839,492	1.12
Arizona.....	5,130,632.....	4,770,861	-7.01	4,807,423	-6.30	4,717,977	-8.04
Utah.....	2,233,169.....	2,194,955	-1.71	2,202,945	-1.35	2,158,202	-3.36
Nevada.....	1,998,257.....	1,856,460	-7.10	1,847,723	-7.53	1,827,109	-8.56
<i>Pacific</i>							
Washington.....	5,894,121.....	5,838,090	-0.95	5,810,813	-1.41	5,941,423	0.80
Oregon.....	3,421,399.....	3,385,241	-1.06	3,384,972	-1.06	3,410,454	-0.32
California.....	33,871,648.....	32,462,610	-4.16	32,377,783	-4.41	33,244,071	-1.85
Alaska.....	626,932.....	650,925	3.83	630,985	0.65	652,081	4.01
Hawaii.....	1,211,537.....	1,253,667	3.48	1,234,930	1.93	1,257,884	3.83
UNITED STATES	281,421,906	274,061,914	-2.62	274,061,954	-2.62	275,462,964	-2.12

Resident Population for April 1, 2000. Series A and B reflect different interstate migration assumptions and do not sum to the same total due to rounding. See text for explanations. Source: Population Division, U.S. Bureau of the Census.

symmetrically distributed to produce an average that reflects more accurately the error represented by most of the observations. The MAPE in most instances is based on a right-skewed, asymmetrical distribution of absolute percentage errors where outliers are likely to pull the summary measure of error upward, thereby overstating the error represented by most of the observations.

Examining data spread. Emerson and Stoto (1983) and Swanson, et al., (2000:194) recommend first looking at the spread of the data to determine if a distribution of APEs appears to be unduly dominated by outliers. They suggest applying a transformation when the ratio of the largest value to the smallest value exceeds 20.

Power transformation. Swanson, et al., (2000) have described a modified Box-Cox power transformation procedure to determine the most symmetrical transformed distribution of APEs, which was defined as:

$$Y = (X^\lambda - \lambda) / \lambda, \text{ for } X \neq 0; \text{ or } Y = \text{LN}(X), \text{ for } X = 0,$$

where X is the untransformed APE, Y is the transformed APE, and λ (lambda) is the power transformation constant. Lambda is determined by finding its value that maximizes the function:

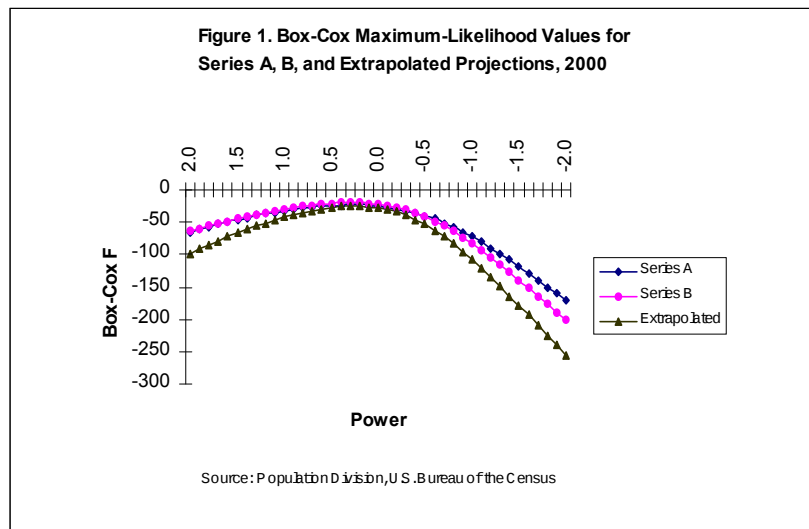
$$\text{ML}(\lambda) = -(N/2)(\text{LN}[(1/N)\sum(Y_i - Y)^2]) + (\lambda - 1)(\sum \text{LN}(X_i)),$$

where N is the number of states, Y_i is the transformed APE, Y is the mean of the transformed APEs, X is the untransformed APE, and $\sum$ represents the sum over all observations. A “coarse grid” search, set up in a Microsoft Excel spreadsheet, was used to solve for values of λ from -2 to 2 inclusive, using increments of $.10$.

Figure 1 shows the nonlinear relationship of the Box-Cox maximum-likelihood values associated with λ for each set of projection APEs. The optimal value of λ (0.3) corresponds to the largest maximum-likelihood value (the smallest negative value in the graph) for Series A, Series B, and extrapolated projection APEs.

Test for skewness. The original APEs and transformed APEs can be compared for skewness using graphic devices and the D’Agostino skewness test. Boxplots

are graphic devices used to identify location, spread,



skewness, tail length, and outliers. The spread of the box represents 50 percent of the values between the first and third quartiles (qt). The boxplot for the untransformed APEs indicates a right-skewed distribution whenever the median (the crossbar in the box) is closer to the lower end of the box with a long upper tail. Similarly, the histogram can be used to visually identify asymmetrical and right skewed distributions (see Figures 2 to 4).

Identifying extreme outliers. Swanson, et al., (2000:196) and Emerson and Strenio (1983:59-60) suggest that extreme population outliers for the original APEs should be mathematically identified using information from the boxplot. They suggest calculating extreme outlier cutoff points by multiplying the fourth spread or width of the middle half of the data by 1.5, adding that product to the third quartile value, and subtracting the resulting sum from the first quartile value.

Calculating MAPE-T and MAPE-R. Once it is established that the transformation of APEs was necessary to correct for skewness and asymmetry, the MAPE-T is calculated from the transformed APE distribution using the APE and MAPE formula discussed above. The next step is to calculate MAPE-R, the re-expressed average that matches the original metric distribution, since it is not easy to interpret MAPE-T, the average of the transformed observations, which is in a different unit of measurement. Swanson, et al., (2000:199) recommended using a nonlinear power function to map the scales of the transformed and original observations such as:

$$X_i = A * Y_i^b,$$

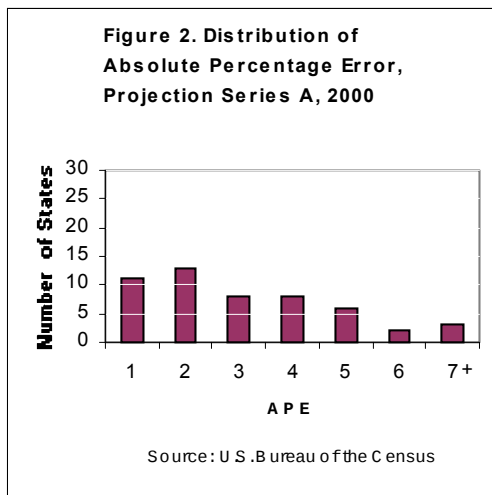
where X is the original APE, Y is the transformed APE, and A and B are estimated parameters. The estimated parameters from the linear regression expressed in logarithm form:

$$\text{LN}(X) = \text{LN}(A) + B [\text{LN}(Y)]$$

can be used with MAPE-T to estimate:

$$\text{MAPE-R} = e^{\{A + (B * [\text{LN}(\text{MAPE-T}))\}}$$

The resulting MAPE-R is reported to be a better measure of the central tendency of the error that is not influenced by the asymmetry and outliers that are found in the original absolute percentage error distribution.

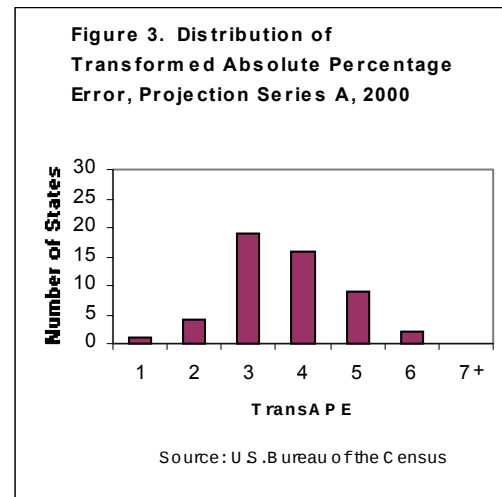


Findings and Conclusions. The findings below suggest the need for a more robust summary measure of forecast error, such as MAPE-R, to evaluate the Census Bureau's state population projections. Additionally, the use of 1) statistical cutoffs to identify extreme outliers, and 2) simple extrapolation as a standard to evaluate state population projections, provides statistical guidelines, rather than subjective conclusions for identifying forecast errors. Data issues associated with comparing the 1990-based projections with the 2000 census are also presented below.

Descriptive analysis. Clearly, the first step in this evaluation was to identify outliers by looking at the percent error for the magnitude and direction of forecast error (see Table 1). There appears to be some overall consistency in the direction of forecast error for states. All three sets of projections underprojected nearly four-fifths of the same states. The few states that were consistently overprojected for all three sets of projections were mostly in the West; i.e., Montana, Idaho, Wyoming, New Mexico, Alaska, and Hawaii,

while the remainder were West Virginia in the South, and South Dakota and North Dakota in the Midwest. For three states, the direction of error was not consistent across all of the projections. Vermont was overprojected only on Series A, while Ohio and Washington were only overprojected on the extrapolated series. The most accurate state projections were those for Alabama on Series A (-0.14 percent) and the extrapolated projections (-0.04 percent), and Ohio on Series B (-0.09 percent).

The range of error was smallest for Series B. The Series B projections ranged from an underprojected population of -7.5 percent for Nevada to an overprojected population of 4.7 percent for Wyoming. In comparison, error in the Series A projections ranged from -8.4



percent for the District of Columbia to 5.8 percent for Wyoming. The range of variation for the extrapolated projections was much wider than either Series A or B, ranging from -11.4 percent for the District of Columbia to 4.0 percent for Alaska.

Among the three sets of population projections, two states, Nevada and Arizona, plus the District of Columbia, consistently stand out as extremely low outliers (see Table 1). Both sets of the Census Bureau projections were more accurate than the extrapolated projection for these three outliers.

Clearly, the simple descriptive review so far suggests that Series B appears to be the most accurate. All three sets of projections had trouble accurately forecasting the up and down swings in population growth that occurred in the West during the 1990's. Additionally, the quality of the 1990 census and post census estimates probably contributes to error in the projections; however, these issues were not the focus of the current study.

MAPEs. A comparison of the MAPEs in Table 2 suggests that Series B is the most accurate for the U.S. and the regions. The MAPE for Series B (2.44 percent) is slightly lower than the extrapolated projections (2.54 percent) and Series A (2.63 percent). Furthermore, Series B tends to underproject the actual population (42 of the 51 states which include the District of Columbia were too low; see Table 1). Nearly half (25 out of 51 APEs) were within 2.5 points of the MAPE for Series B.

Forecast error for the regions and divisions varied greatly, but tend to be consistent across the three sets of projections. The MAPEs for regions were highest for the West and lowest for the Midwest. Two-thirds of the division MAPEs were lower for the extrapolated projections than for the Series A and B projections. Due to the smaller number of observations for MAPEs at the region and division level, no attempt was made to validate the region or division results. The next step is to measure the variation in the APEs and determine if they are asymmetrically distributed.

Spread and Asymmetry. Review of the data found a wide range of variation in the APEs which warrants the application of the Emerson-Stoto spread ratio to the original distribution. The transformation of the original distribution of errors was performed since the spread ratio of the highest original APE to the lowest original APE exceeds 20 for each of the three projections. In Series A, the highest APE, 8.38, is for the District of Columbia, while the lowest APE, 0.14, is for Alabama which results in a spread ratio of 60 (8.38/0.14). For Series B, the highest APE, 7.53 in Nevada, and the lowest APE, 0.09 in Ohio, results in a ratio of 84. The widest range occurs in the extrapolated projections where the highest APE, 11.44 for the District of Columbia, and the lowest APE, 0.04, for Alabama, results in spread ratio of 286.

In contrast, transformed APEs for Series A and B, and the extrapolated projections had Emerson-Stoto spread ratios below 20. For example, the log-percentage errors for the transformed APEs (not shown) for Series A ranged from a high of 5.31 percent for the District of Columbia to a low of 0.83 percent for Alabama, which results in a spread ratio of 6. Similarly, the transformed APEs for Series B ranged from 5.11 percent for Nevada to 0.60 percent for Ohio, with a spread ratio of 9. The transformed APEs for the extrapolated projection ranged from a high of 5.92 percent for the District of Columbia to a low of 0.31 for Alabama, with a spread ratio of 19.

After calculating transformed MAPEs (MAPE-T) using

the modified Box-Cox method, histograms and boxplots are created to evaluate the shape of both the original and the transformed distributions. Histograms of the original APEs show data that are asymmetrical and slightly right-skewed (see Series A in Figure 2), while

Table 2. Mean Absolute Percentage Error in State Population Projections, By Region And Division From Series A and B, And Extrapolated Projections, 2000

Region and division	Series A	Series B	Extrapolation
United States	2.63	2.44	2.54
Northeast	2.50	2.57	3.15
New England	2.42	2.58	3.51
Middle Atlantic	2.67	2.55	2.43
Midwest	1.58	1.40	1.11
East North Central	1.54	1.36	1.07
West North Central	1.60	1.43	1.14
South	2.60	2.58	2.69
South Atlantic	3.50	3.50	3.90
East South Central	0.89	0.87	0.86
West South Central	2.29	2.21	1.79
West	3.75	3.13	3.22
Mountain	4.41	3.91	3.88
Pacific	2.69	1.89	2.16

Mean absolute percentage error (MAPEs) are results for 5 years-out from the 1995 population. Based on the enumerated 2000 census counts, the 2000 population for Series A, B, and Extrapolated Projections derived from the Absolute Percentage Errors calculated for the states and the District of Columbia, see text for detailed explanation. Source: U.S. Bureau of the Census.

the histograms for the transformed APEs (TransAPEs) are symmetrical (see Series A in Figure 3). The same patterns were found for Series B and the extrapolated projections, but the histograms are not shown here. The histograms for the three sets of projections validated the need for the data transformations.⁷

Similarly, the boxplots in Figure 4 confirm that original APEs are asymmetrical and right-skewed for all three sets of projections. The median (the crossbar in the box) appears between the middle and bottom of the box, with a long upper tail for the three original APE distributions. The box spread is narrower for the transformed APEs (TransAPE or T-APE in the graph) and symmetrical with the median in the middle of the box (the same location as the mean), with a lower and upper tail of equal length. All of the calculations and graphs were derived using Microsoft Excel, which does not easily facilitate showing the asterisks for extreme outliers in the boxplot graphs.

⁷ The histogram sorts the APE and TransAPE distributions in a Microsoft Excel spreadsheet using bin range values of 1 to 7. The lowest bin values include APE or TransAPE values less than one, while the highest values were included with 7. A rule of thumb for choosing the APE bin range is one minus the highest value and one plus the lowest value.

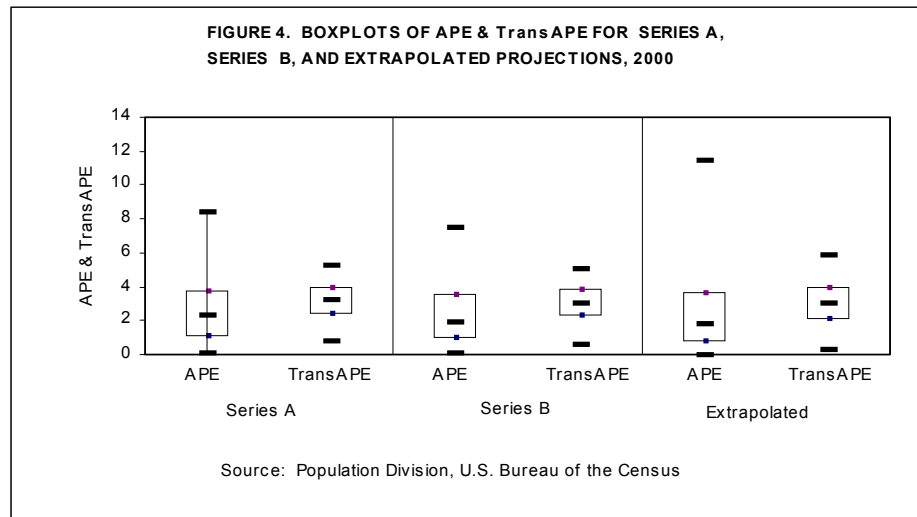
The skewness coefficients of 1.02 for Series A, 1.05 for Series B, and 1.74 for the extrapolated projections imply that each set of projections was asymmetrical and right-skewed. A symmetrical distribution would have a skewness coefficient of zero. The D'Agostino skewness test suggests that the null hypothesis (the data are not skewed) should be rejected ($p = 0.000$) for all three of the original APE distributions.

Extreme outliers. The Emerson-Strenio “fourth spread” procedure was used to identify extreme outliers among the original APE distributions. The “fourth spread” upper cutoff values were 7.80 for Series A, 7.38 for Series B, and 7.93 for the extrapolated projections. The District of Columbia and Nevada were identified as the only extreme outlier in Series A and B, respectively. In the extrapolated projections, three extreme outliers with APEs above the cutoff value were Arizona, the District of Columbia, and Nevada.

MAPE-R results. While the individual transformed APEs are of no interest in the evaluation of the state projections, the summary statistic for the transformed APEs is useful. Additionally, the MAPE-T which equals 3.21, 3.10, and 3.04 for Series A, B, and the extrapolated projections, respectively, is difficult to explain, since the results are log-based. In order to explain the transformed summary statistic in the original metric format, the next step is to re-express MAPE-T as MAPE-R using the logarithm regression results (see formula discussed earlier).⁸ The MAPE-R derived for Series B projections at 2.06 percent is slightly more accurate than Series A at 2.24 percent. Additionally, the MAPE-R for the extrapolated projection at 2.00 percent was about the same as the Series B projections and slightly more accurate than the Series A projections.

Table 3 shows MAPE overstating the forecast error in

comparison to the MAPE-R. The ratio of the MAPE to the median (absolute percent error) is another useful descriptive tool that shows the overstated forecast error (Tayman and Swanson 1999:307). In Table 3, the



MAPE-to-median ratios confirm that MAPE overstates forecast error, since the ratios are greater than 1.0 for all three projections. A different conclusion would have been drawn if the original error distribution were not corrected for skewness and asymmetry.

Initially, Theil's U was considered as a potential summary measure to determine if the Census Bureau's forecast models were more accurate than the extrapolated projections. However, it was not accepted as a valid measure since the distributions of APEs were found to be skewed and asymmetrical. Armstrong and Collopy (1992:77) reported that RMSE (used to derive Theil's U) is unreliable due to its poor protection against outliers.⁹ Additional issues related to the guidelines used for choosing appropriate forecast error measures are discussed by Ahlburg (1992).

To summarize, MAPE-R was used to replace the summary measure MAPE in the evaluation of the Census Bureau's projections since the data distributions were skewed and asymmetrical. The results show that the Census Bureau's state population projections for April 2000 (Series B - the economic model) had the least forecast error, with an average absolute percentage

⁸ The logarithm regression results used to derive MAPE-R were (1) Series A: $A = -1.950$, $B = 2.366$, $R^2 = 0.992$, standard error (SE) = 0.074; (2) Series B: $A = -1.850$, $B = 2.275$, $R^2 = 0.986$, and SE = 0.104 and (3) the extrapolated projections: $A = -1.649$, $B = 2.104$, $R^2 = 0.968$, and SE = 0.194.

⁹ Theil's U, interpreted as the RMSE of the projections model divided by the RMSE of the extrapolated or no-change model, is derived from the formula: $U = [\sum (P_i - E_i)^2]^{1/2} / [\sum E_i^2]^{1/2}$, where P_i refers to the projection error for each state, and E_i is the corresponding extrapolation error used as the standard for each state (see Kolb and Stekler, 1993). Using either absolute or percent change or the transformed percent change, Theil's U coefficients found both Series A and B with fewer errors than the extrapolated projections.

error of 2.06 percent. This is slightly better than Series A, with an average absolute percentage error of 2.24 percent. The forecast error in the Census Bureau's Series B projections was the same as that found in the extrapolated projections (2.00 percent), while the extrapolated projections slightly out-performed Series A - the preferred series. All three projections consistently underprojected approximately four-fifths of the states (40 states in the extrapolated projections, 41 states in Series A, and 42 states in Series B) out of a total of 51 states (including the District of Columbia). The widest range of variation and the most extreme outliers were found for the extrapolated projections.

Table 3. Comparison of the MAPE, Ratio to Median, and MAPE-R for Series A, Series B, and Extrapolated Projections, 2000.

	Ratio of MAPE to Median		MAPE-R
Series A	2.63	1.13	2.24
Series B	2.44	1.24	2.06
Extrapolated	2.54	1.38	2.00

Summary Statistics for projections evaluation using enumerated Census 2000 results, see text for detailed explanation.
Source: Population Division, U.S. Bureau of Census.

An added feature of the extrapolated projection is that base period (1990-1995) growth trends are held constant over the projection horizon (1995-2000). This information is useful for identifying changes in trends (or error) between the base period and the projection horizon. Ideally, when the extrapolated projection error is zero, there is no evidence of change in the pattern of growth between the base period and the projection horizon. In this study, nearly a third of the states (16 states with error ranging from 1.0 percent to -1.0 percent) showed little change in the 1990-95 pattern of population growth extrapolated to 2000.

Several issues or differences between the 1990 and 2000 censuses not examined in this study probably affect the accuracy of the state projections. First, adjusted 1990 census counts were not used as the base year and any undercoverage in the 1990 census is carried throughout the post-1990 estimates and projections.¹⁰ Second, this evaluation only examines the aggregated population totals and does not evaluate the separate component totals, such as births, deaths,

state-to-state, and international migration, by age, sex, and race/Hispanic origin. The domestic migration and international migration components are the most difficult to adequately baseline or project. Additionally, retrospective census information on place of residence during 1985-90 used in the projections may not reflect changes in the age pattern of migrants during the 1990's. Third, the race/Hispanic origin categories are quite differently defined in each of the censuses, the vital statistics, and administrative records. Fourth, the state projections use national data as a proxy in the absence of detailed demographic components. Mulder (2001) evaluating the Census Bureau's national population projections produced between 1947 and 1994 has documented the inability of past projections to accurately forecast turning points, particularly for the immigration and fertility components of the projections. Finally, there is the issue of the multi-dimensional raking, in other words the state projection results are aggregated pro-rata to the national estimates and projections for consistency at the national level by age, sex, and race/Hispanic origin.

The 2000 state population projections appear to be slightly more accurate than vintage projections produced decades earlier. Wetrogan and Campbell (1990) calculated MAPEs ranging from 3.0 percent to 5.2 percent for a 5-year projection horizon in their evaluations of 1970's and 1980's Census Bureau projections using corresponding 1970's and post-1980 census estimates.¹¹ They reported U.S. MAPEs from the Census Bureau's 1987 projection at 0.5 percent, 1.1 percent, and 1.6 percent for one-, two-, and three-year projection horizons, respectively. MAPEs of 0.5 percent per year appear to be a reasonable level of accuracy to expect for state population projections over a short term or 5-year projection horizon.

This study found that the Census Bureau's 2000 state population projections are as accurate as simple extrapolated projections and have fewer extreme outliers. Further evaluation of the detailed demographic components should aid in identifying areas of the projection model that needs to be improved. It appears that tests for skewness and asymmetry are necessary to validate the use of the popular summary measure, such as the MAPE or its variant MAPE-R.

The advantage of using MAPE-R in conjunction with the original absolute percentage error is that users are more familiar with interpreting this summary measure and MAPE-R resolves the central tendency issues

¹⁰ An evaluation of the factors affecting the accuracy of the state projections, such as census undercount, estimates error, and error in the projected components of change have been addressed by Wang (2002).

¹¹ An evaluation of the 1970's and 1980's state population projections using the 1990 post census estimates, final intercensal estimates, and MAPE-R would probably yield lower forecast errors.

whenever MAPE is found to be invalid. Clearly, a drawback to its widespread use is the cumbersome statistical calculation needed to carry out its application; nevertheless, all of the results for this evaluation were carried out in Microsoft Excel spreadsheets.¹² With a few modifications, the spreadsheets can be used to evaluate error in other small subnational estimates or projection data sets.

Acknowledgments. I would like to thank Greg Spencer and Edwin Byerly for inspiring this research, and Ching-Li Wang for his comments and sharing evaluation results.

References

- Ahlburg, D.A. 1992. "A Commentary on Error Measures: Error Measures and the Choice of a Forecast Method." *International Journal of Forecasting* 8:99-111.
- Armstrong, J.S. 2001. "The Forecasting Dictionary" in *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, M.A., Kluwer Academic Publishers.
- Armstrong, J.S. and F. Collopy. 1992. "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons." *International Journal of Forecasting* 8:69-80.
- Armstrong, J.S. 1977. *Long-range forecasting, from Crystal Ball to Computer*. New York: John Wiley & Sons.
- Box, G. and D. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society. Series B* 26:211-52.
- Campbell, P.R. 1996. "Population Projections for States, by Age, Sex, Race, and Hispanic Origin: 1995 to 2025." PPL-47. U.S. Bureau of the Census, Population Division. Data files: PE-45 on Internet <http://www.census.gov/population/www/projections/stproj.html>.
- Campbell, P.R. 1997. "An Evaluation of the Census Bureau's 1995 to 2025 State Population Projections – One Year Later." Presented at the Population Association of America Meeting: Washington, DC.
- Coleman, C.D. 2000. "Evaluation and Optimization of Population Projections Using Loss Function." in *Papers and Proceedings of the 11th Federal Forecasters Conference - 2000*. edited by D. E. Gerald. U.S. Department of Education, Office of Educational Research and Improvement. Washington, D.C.
- Coleman, C.D. 2002. "Measures of Estimates Quality." Presented at the Population Association of America Meeting. Atlanta, GA.
- D'Agostino, R.A. et. al. 1990. "A Suggestion for Using Powerful and Informative Tests of Normality." *The American Statistician*. 44:316-21.
- Davis, S.T. 1994. "Evaluation of Postcensal County Estimates for the 1980s." Working Paper 5. U.S. Bureau of the Census, Population Division.
- Day, J.C. 1996. *Population Projections of the United States by Age, Sex, Race, and Hispanic Origin: 1995 to 2050*. Current Population Reports. Series P25-1130. U.S. Bureau of the Census. Washington, D.C. Government Printing Office.
- Emerson, J. and M. Stoto. 1983. "Transforming Data." Pp. 97-128 in *Understanding Robust and Exploratory Data Analysis*, edited by D. Hoaglin, F. Mosteller, and J. Tukey. New York: Wiley.
- Emerson, J. and J. Strenio. 1983. "Boxplots and Batch Comparisons." Pp. 58-96 in *Understanding Robust and Exploratory Data Analysis*, edited by D. Hoaglin, F. Mosteller, and J. Tukey. New York: Wiley.
- Kolbs, R.A. and H.O. Stekler. 1993. "Are Economic Forecasts Significantly Better than Naïve Predictions? An Appropriate Test." *International Journal of Forecasting* 9:117-120.
- Mulder, T. 2000. "Accuracy of the U.S. Census Bureau National Population Projections and Their Respective Components of Change." In *Papers and Proceedings of the 11th Federal Forecasters Conference - 2000*. edited by D. E. Gerald. U.S. Department of Education, Office of Educational Research and Improvement. Washington, D.C.
- Shryock, H.S. and J.S. Siegel. 1976. *The Methods and Materials of Demography*. Condensed Edition. San Diego, California: Academic Press.
- Smith, S.K. and T. Sincich. 1992. "Forecasting State and Household Populations, Evaluating the Forecast Accuracy and Bias of Alternative Population Projections for States." *International Journal of Forecasting* 8: 495-508.
- Snedecor, G.W. and W.G. Cochran. 1980. *Statistical Methods*. Ames: Iowa State University Press.
- Swanson, D.A., J. Tayman, and C.F. Barr. 2000. "A Note on the Measurement of Accuracy for the Subnational Demographic Estimates." *Demography* 37:193-201.
- Tayman, J. and D.A. Swanson. 1999. "On the Validity of MAPE as a Measure of Population Forecast Accuracy." *Population Research and Policy Review* 18:299-322.
- U.S. Bureau of the Census. 2000. "Table 2. Resident Population of the 50 States, the District of Columbia, and Puerto Rico: Census 2000." December 28. Internet site: <http://blue.census.gov/population/cen2000/tab02.pdf>.

¹² A copy of the Microsoft Excel spreadsheet used to evaluate the state projections can be obtained from the author by e-mail addressed to Paul.R.Campbell@census.gov.

U.S. Bureau of the Census. 1996a. "Estimates of Population and Demographic Components of Change for States: Annual Time Series, 1990-96." ST-96-1. Population Distribution Branch.

U.S. Bureau of the Census. 1996b. "Population of States by Single Years of Age and Sex: 1990 to 1995." PE-38. Population Division.

U.S. Bureau of the Census. 1996c. "Estimates of the

Population of States by Age, Sex, Race, and Hispanic Origin: 1990 to 1994." PE-47. Population Division.

Wetrogan, S.I. and P.R. Campbell. 1990. "Evaluation of State Population Projections." Presented at the Population Association of America: Toronto, Canada.

Wang, C. 2002. "Evaluation of Census Bureau's 1995-2025 State Population Projections" Population Projections Branch. U.S. Bureau of the Census, Population Division. Forthcoming.