Prediction Intervals for County Population Forecasts

Stefan Rayer, University of Florida
Stanley K. Smith, University of Florida
Jeff Tayman, San Diego Association of Governments

Paper presented at the annual meeting of the British Society for Population Studies, Southampton, UK, September 18-20, 2006

ABSTRACT

Population forecasts entail a significant amount of uncertainty, especially for long-range horizons and for places with small or rapidly changing populations. This uncertainty can be dealt with by presenting a range of projections or by developing statistical prediction intervals based on models that incorporate the stochastic nature of the forecasting process or on empirical analyses of past forecast errors. In this paper, we develop and test empirical prediction intervals for county population forecasts in the United States. We find that prediction intervals based on the distribution of past forecast errors provide reasonably accurate predictions of the distribution of future forecast errors. We believe the construction of empirical prediction intervals to accompany population forecasts will help data users plan more effectively for an uncertain future.

Introduction

Population forecasts entail a significant amount of uncertainty, especially for long-range horizons and for places with small or rapidly changing populations. More than 30 years ago, Keyfitz (1972) made the case that demographers should provide a warning regarding that uncertainty to the users of their forecasts. This warning has typically been provided by presenting a range of projections (e.g., Hollmann, Mulder, & Kallan, 2000), but in recent years attention has been given to developing statistical prediction intervals that provide an explicit probabilistic statement regarding the level of error expected to accompany a population forecast. Statistical prediction intervals can be based on models that incorporate the stochastic nature of the forecasting process (e.g., Alho and Spencer, 1990; Cohen, 1986; Lutz, Sanderson, and Scherbov, 1999; Pflaumer, 1992) or on empirical analyses of past forecast errors (e.g., Keyfitz, 1981; Smith and Sincich, 1988; Stoto, 1983; Tayman, Schafer, and Carter, 1998).

In this paper, we develop and test prediction intervals based on the latter approach. Under formal definitions, probability statements regarding the accuracy of population forecasts based on the distribution of past forecast errors cannot be made because the distribution of future errors is unknown (and unknowable) at the time the forecasts are made. However, if current forecasting methods 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, then we can assume that future forecast errors will be drawn from the same distribution as past forecast errors (Keyfitz, 1981). If this is true, empirical prediction intervals will provide a reasonable measure of the uncertainty surrounding current population forecasts.

The usefulness of empirical prediction intervals relies heavily on the assumption that the distribution of forecast errors remains stable over time. Few researchers have evaluated the validity of this assumption. Perhaps the most comprehensive evaluation was conducted by Smith and Sincich (1988), who examined state-level population forecasts using data from 1900 to 1980. Following a methodology developed by Williams and Goodman (1971), they evaluated forecast errors for 10- and 20-year horizons and found that the means and variances of absolute forecast errors remained relatively stable over time, especially after 1920, and that the variances of algebraic forecast errors remained moderately stable over time but their means were not at all stable. They concluded that the study of past forecast errors is useful for forecasting the level of precision of current population forecasts, but not for forecasting their tendency to be too high or too low.

Since that study, little additional research has analyzed the stability of forecast errors over time or investigated the performance of empirical prediction intervals. To our knowledge, no study has considered these issues at the substate level. We believe research at the substate level is essential because small-area forecasts are used by decision makers for a wide variety of planning, budgeting, and analytical purposes. Examples include planning for future water consumption (Texas Water Development Board, 1997), choosing locations for new fire stations (Tayman, Parrott, and Carnevale, 1994), evaluating the demand for additional hospital services (Thomas, 1994), and projecting future public school enrollment (McKibben, 1996). Optimal decisions cannot be made without some understanding of the likely level of accuracy of the population forecasts upon which those decisions are based.

In this paper, we analyze population forecast errors for counties in the United States. Following the approach used by Smith and Sincich (1988), we construct empirical prediction intervals and investigate whether error distributions from previous forecasts provide useful predictions of error distributions for subsequent forecasts. We do not conduct formal statistical tests, but rather evaluate stability indirectly using averages, medians, 90^{th} percentile errors, and coefficients of variation. We find that the study of past forecast errors can indeed provide useful information regarding the likely distribution of future forecast errors. We believe this information provides a basis for constructing empirical prediction intervals that will help data users evaluate the likely accuracy of population forecasts and plan more effectively for an uncertain future.

Data and Forecasting Techniques

We used decennial census data from 1900 to 2000 to construct and analyze population forecasts for counties (or county equivalents) in the United States.[1] We restricted our analysis to the 2,482 counties for which there were no significant boundary changes between 1900 and 2000; this group accounted for 79% of all counties in 2000. Forecast errors for this group of counties were compared to forecast errors for a larger group of 2,978 counties (accounting for 95% of all counties) whose boundaries did not change significantly after 1930. Precision and bias for these two groups of forecasts were found to be very similar. We used the smaller group with constant boundaries since 1900 because it permitted the analysis of a larger number of launch years and forecast horizons.

We use the following terminology to describe population forecasts:

1) Base year: the year of the earliest population size used to make a forecast.

2) Launch year: the year of the latest population size used to make a forecast.

3) Target year: the year for which population size is forecasted.

4) Base period: the interval between the base year and launch year.

5) Forecast horizon: the interval between the launch year and target year.

For example, if data from 1900 and 1920 were used to forecast population in 1930, then 1900 would be the base year, 1920 would be the launch year, 1930 would be the target year, 1900 - 1920 would be the base period, and 1920 - 1930 would be the forecast horizon.

We made forecasts of total population for each county using seven simple trend extrapolation techniques (see Appendix 1 for a description of these techniques). The forecasts were based on 20-year base periods, the base period shown previously to produce the most accurate forecasts for counties in this data set (Rayer, 2004). The forecasts had launch years extending from 1920 to 1990 and horizons ranging from 10 to 30 years. The 21 combinations of launch year and forecast horizon—and their associated target years—are shown in Table 1.

Compared to other techniques, simple trend extrapolation techniques have a number of advantages for population forecasting purposes. They require few base data, can be applied at low cost, and can be applied retrospectively to produce forecasts that are comparable over time. These characteristics are particularly important when making forecasts for a large number of geographic areas and historical time periods. Furthermore, a substantial body of evidence indicates that trend extrapolation techniques produce forecasts of total population that are at least as accurate as those produced by more complex techniques (for a summary of the evidence, see Smith, Tayman, and Swanson, 2001: 307-313). We believe these techniques provide a useful vehicle for assessing the stability of population forecast errors over time and testing the validity of empirical prediction intervals.

We calculated the average of the seven individual forecasts for each county (AV7) and the average after the highest and lowest were excluded (AV5). The latter measure reduces the impact of outliers on forecast errors and is often called a trimmed mean; we found it produced slightly smaller forecast errors than AV7 in the present study. A number of studies have documented the benefits of combining forecasts, both in demography and other fields (Armstrong, 2001: 417-439; Smith et al., 2001:328-331). Given the large number of individual forecasts, we present only the results for AV5; however, it should be noted that many of the results for the other techniques were similar to those reported here.

Forecasts for each county were made for each of the 21 launch year/forecast horizon combinations shown in Table 1 and were compared to census counts for each target year. The resulting differences are called forecast errors, although they may have been caused partly by errors in the census counts themselves. All errors are reported as percentages by dividing by census counts and multiplying by 100. We refer to errors that ignore the direction of the error as absolute percent errors (APEs) and errors that account for the direction of the error as algebraic percent errors (ALPEs).

General Description of Forecast Errors

Several summary measures were used to provide a general description of forecast errors. The mean absolute percent error (MAPE), median absolute percent error (MEDAPE), and 90th percentile error (90thPE, calculated as the APE that was larger than exactly 90% of all APEs) are measures of precision; they show how close the forecasts were to population counts regardless of whether they were too high or too low. The mean algebraic percent error (MALPE) and median algebraic percent error (MEDALPE) are measures of bias; they show the tendency for forecasts to be too high or too low. These and similar measures have often been used to evaluate the accuracy of population forecasts (e.g., Isserman, 1977; Pflaumer, 1992; Rayer, 2004; Smith and Sincich, 1988; Tayman et al., 1998).

We also used a measure of the distribution of APEs. The coefficient of variation (CV) is the standard deviation divided by the mean error, multiplied by 100. It provides a measure of the dispersion of forecast errors around the mean value. Theoretically, CV values can range from zero to infinity. In reality, they are seldom zero and rarely approach infinity, although they are often found to be greater than 100. High values reflect a high degree of dispersion around the mean and low values reflect a low degree of dispersion. When measured over time, high CV values reflect a low degree of temporal stability and low values reflect a high degree of stability. The CV provides a way to compare the degree of dispersion in one data series with that in another, even if the means differ substantially from each other.[2]

How can CV values be judged? That is, what values reflect high, medium, or low degrees of stability? There are no clear guidelines in the literature; rather, values have been found to differ substantially from one context to another, depending on the specific variables, geographic regions, and time periods covered. For example, CV values for measures of athletic performance have been found to fall between 1% and 5%, depending on the nature of the event, the time between events, and the experience of the athlete (Hopkins, 2000). Interest rates for household saving deposits have exhibited CV values between 8% and 28% for countries in the European Union (European Central Bank, 2006). A study of the commercial television industry found a CV of 94% in turnover rates for managers (Sorenson, 2002). A study of subcounty forecast errors in San Diego County found CV values ranging from 75% to 235% for populations in different size categories (Tayman, et al., 1998). In analyzing the stability of forecast errors over time, we classify CV values of less than 10% as very stable, 10-25% as stable, and greater than 25% as unstable.

Table 2 shows forecast errors for counties by target year and forecast horizon. Several patterns stand out in the measures based on APEs. First, the MAPE exceeded the MEDAPE for every horizon and target year, indicating that even though the trimmed mean (AV5) excluded the individual forecasts with the highest and lowest values, MAPEs were still affected by the presence of outliers. Typically, MAPEs exceeded MEDAPEs by 30–40%. Second, errors increased about linearly with the forecast horizon. For each ten year increase in the forecast horizon, MAPEs rose by about 10%, MEDAPEs by about 7%, and 90thPEs by about 21%. Third, there were only modest differences in errors by target year within each forecast horizon, at least until the last few target years. Only for 1990 and 2000 for 10-year horizons and 2000 for the 20-year horizon were errors substantially different than for all other target years. We offer an explanation for this finding later in the paper.

Whereas MAPEs always exceeded MEDAPEs, MALPEs were sometimes larger than MEDALPEs and sometimes smaller. Both of these measures of bias varied considerably by target year, as is shown by the large standard deviations and changes in sign from one target year to another. However, there was a tendency for MALPEs and MEDALPEs to be positive for earlier target years and negative for later target years, especially for longer forecasts horizons. In general, we believe that there is no predictability regarding the likelihood that a given set of forecasts will turn out to be too high or too low. A number of previous studies have drawn similar conclusions (e.g., Isserman, 1977; Kale, Voss, Palit, and Krebs, 1981; Smith and Sincich, 1988; Tayman et al., 1998).

The CV generally declined as the target year increased, although the differences from one target year to another were not extremely large. This implies that APEs became slightly more concentrated around the mean as the century progressed. It is particularly noteworthy that—in contrast to the other measures—CVs changed very little as the forecast horizon grew longer. This shows that even though the MAPE, MEDAPE, and 90thPE increased steadily with the length of the forecast horizon, the degree of dispersion of APEs around the mean remained quite stable.

The bottom three rows of each panel in Table 2 provide a summary of the results for all the individual target years within a given length of forecast horizon; that is, the average, standard deviation, and CV are based on the values of each error measure for each target year. The CV in the bottom row has a different interpretation than the CV discussed above. Whereas the CV for the right-hand column of the table shows the degree of dispersion of APEs around the mean for *individual* target years, the CV in the bottom row of each panel shows the degree of dispersion of MAPEs, MEDAPEs, and 90^{th}PEs *across* target years.

Although the averages varied considerably from measure to measure, the CVs for the MAPE, MEDAPE, and 90^{th}PE were very similar to each other within each of the three horizons, ranging only from 22 to 25 for 10-year horizons, from 20 to 22 for 20-year horizons, and from 13 to 17 for 30-year horizons. The CVs for the CV had even lower values. These results reflect a fairly high degree of stability in the distribution of APEs over time and provide support for the hypothesis that empirical prediction intervals based on errors from one time period will provide reasonable measures of uncertainty for subsequent population forecasts.

Empirical Prediction Intervals

Although the data in Table 2 show a substantial degree of stability over time in the distribution of APEs, they show no stability at all in the distribution of ALPEs. This suggests that the study of past forecast errors may help us predict the level of precision of current forecasts, but is not likely to help us predict their tendency to be too high or too low. We therefore focus on the distribution of APEs in our efforts to develop and evaluate empirical prediction intervals.

Under the approach used by Smith and Sincich (1988), information on the distribution of past APEs is used to predict the distribution of future APEs. A major advantage of this approach is that it can accommodate any type of error distribution, including the asymmetric and truncated distributions characteristic of APEs. It also permits an assessment of the prediction intervals themselves; that is, we can compare the actual number of errors falling within the intervals with the predicted number.

Following this approach, we ranked the APEs for each of the 21 sets of forecasts and determined the 90^{th}PE, as shown in Table 2. Then, we used the 90^{th}PE from target year *t-n* as the forecast of the 90^{th}PE in target year *t*, where *n* is the length of the forecast horizon. For example, if 1950 were the target year for a 10-year forecast based on launch year 1940, the 90^{th}PE for 1950 would be used to predict the 90^{th}PE for 1960 for a 10-year forecast based on launch year 1950. If error distributions remain relatively stable over time, 90^{th}PEs from past distributions will provide reasonably accurate predictions of future 90^{th}PEs. To assess the validity of that hypothesis, we compared the predicted with the actual 90^{th}PE for each target year and computed the percentage of APEs that fell within the predicted values.

Table 3 shows the percentage of APEs in each target year that was less than the predicted 90^{th}PE. The numbers can be interpreted as follows: A value of 90 reflects a perfect prediction. Values below 90 indicate that the 90^{th}PE for target year *t* was greater than the 90^{th}PE for target year *t* – n (i.e., fewer APEs fell within the predicted range). Values above 90 indicate the opposite. In addition to errors for each target year, this table shows 90^{th}PEs averaged across all target years for each horizon, along with the standard deviation and CV associated with the average 90^{th}PE.

For averages covering all the target years within a given forecast horizon, Table 3 reflects a high degree of stability for horizons of differing lengths: approximately 91% of APEs fell within the predicted 90^{th}PE for all three horizons. There was more variability when comparing individual target years within each horizon, but for the most part the values did not stray far from 90, indicating a reasonably high degree of stability over time. CVs were slightly above 6% for all three horizons, further demonstrating temporal stability.

It is possible that using data from several historical time periods to predict future forecast errors will provide better results than using data from a single time period. To test this hypothesis, we evaluated the percentage of 90^{th}PEs that were less than the average of the *two* previous target years (not shown here). This adjustment had little impact on the results, generally leading to errors that were slightly *larger* than those shown here. In this sample, then, data from a single time period were sufficient for constructing empirical prediction intervals.

In order to investigate the impact of the choice of cut-off points for the prediction intervals, we replicated the analysis using 75^{th} percentile errors (75^{th}PE) instead of 90^{th} percentile errors (not shown here). This led to generally similar results, albeit with somewhat more variability from one target year to another. We believe this greater variability was caused by the greater concentration of APEs around the 75^{th}PE than the 90^{th}PE. As a result, small differences in the size of the predicted percentile error led to a larger difference in the percentage of APEs falling within the predicted value for the 75^{th}PE than for the 90^{th}PE. In general, the further the distance from the center of an error distribution, the lower the concentration of APEs around a particular percentile error.

Many studies have found forecast errors to be affected by differences in population size and growth rate. Table 4 shows 90^{th}PEs for counties by population size in the launch year, for each combination of target year and forecast horizon. For each target year and length of horizon, errors generally declined as population size increased, with the largest declines typically occurring in the move from the smallest to the next-smallest size category. The CVs also generally declined as population size increased, reflecting a higher degree of year-to-year variation in errors for small counties than large counties. Many studies have found forecast errors to be larger for small places than large places (e.g., Isserman, 1977; Murdock, et al. 1984; Rayer, 2004; Smith et al., 2001; White, 1954).

How do differences in population size affect stability in the distribution of forecast errors over time? Table 5 shows the percentage of APEs that were less than the predicted 90thPE by population size in the launch year. In general, differences by population size were fairly small and followed no consistent pattern. For some combinations of target year and length of horizon, the percentages rose with population size; for others, they fell; and for some, they followed no clear pattern. The CVs were small and did not vary much among the four size categories or by length of forecast horizon. Although 90thPEs themselves varied considerably with differences in population size, it appears that differences in population size had no consistent impact on the predictability of 90thPEs.

Table 6 shows 90thPEs for counties by the rate of population growth during the base period, for each combination of target year and forecast horizon. Errors generally displayed a U-shaped pattern, with higher values for counties with large negative growth rates, smaller values for counties with moderate growth rates, and higher values for counties with large positive growth rates. These patterns are also consistent with those found in previous research (e.g., Isserman, 1977; Murdock et al, 1984; Smith, 1987). CVs followed the same U-shaped pattern for 10- and 20-year forecast horizons, but followed a continuously upward-sloping pattern for the 30-year horizon.

Table 7 shows the percentage of APEs that were less than the predicted 90thPE by population growth rate during the base period. In contrast to differences in population size, differences in growth rates had a generally consistent impact on the performance of prediction intervals: there was a strong tendency for the percentage of APEs that was less than the predicted 90thPE to increase with the growth rate. Values were generally smallest for counties in the lowest growth category and increased with increases in the growth rate. Furthermore—as indicated by CVs that declined as growth rates increased for all three lengths of forecast horizon—values for individual target years varied most for counties with rapidly declining populations and varied least for counties with rapidly growing populations. That is, there was more consistency in the results across target years for rapidly growing populations than for rapidly declining populations.

These results suggest that differences in the rate of population growth had a consistent impact on the stability of the distribution of forecast errors over time. For counties with rapidly declining populations, there was a tendency for the error distribution from the previous target year to under-project the 90thPE; whereas for counties with rapidly growing populations, there was a tendency to over-project the 90thPE.

The results shown in Tables 4 and 6 provide an explanation for why MAPEs, MEDAPEs, and 90thPEs were smaller for target years at the end of the 20th century than for target years earlier in the century, as shown in Table 2. Over the course of the century, the number of large counties increased and the number of small counties declined; similarly, the number of counties with moderate growth rates increased relative to the number with extreme growth rates. Both of these trends raised the number of counties that tend to have relatively small forecast errors and lowered the number that tend to have relatively large forecast errors. Consequently, errors were smaller for the last years in the century than for earlier years.

Summary and Conclusions

In this study, we evaluated population forecast errors for 2,482 counties in the United States. The forecasts were made using seven simple trend extrapolation techniques and a variety of base periods and forecast horizons between 1900 and 2000. We found that:

1) MAPEs, MEDAPEs, and 90thPEs remained fairly constant over time, but declined over the last few decades in the century.
2) MALPEs and MEDALPEs did not remain at all constant over time.
3) MAPEs, MEDAPEs, and 90thPEs increased with the length of the forecast horizon, often in a nearly linear manner.
4) In most instances, the 90thPE from one time period provided a reasonably accurate forecast of the percentage of APEs falling within the predicted 90% interval in the following time period, even for long forecast horizons.
5) Differences in population size had little impact on the percentage of APEs falling within the predicted 90% interval, but differences in population growth rate had a substantial impact.

Based on this evidence, we have concluded that the study of past forecast errors can provide useful information regarding the distribution of future APEs, but can provide little guidance regarding the tendency for forecasts to be too high or too low. Of particular interest is the finding that—throughout the 20th century— 90thPEs from previous error distributions provided reasonably accurate predictions of subsequent 90thPEs. Given the tremendous changes in population trends that occurred during the 20th century, this is quite a notable finding. It suggests that data from previous time periods can be used to construct empirical prediction intervals to accompany county population forecasts, and that these intervals are likely to provide data users with a realistic measure of the degree of uncertainty inherent in population forecasts that will enhance their ability to plan intelligently for the future.

As this paper shows, there is a substantial degree of uncertainty inherent in short-range county population forecasts and an even higher degree of uncertainty in long-range forecasts. Approximately 10% of the absolute errors in our analysis were greater than 22% for 10-year horizons, greater than 41% for 20-year horizons, and greater than 63% for 30-year horizons. In addition, some sets of forecasts had an upward bias and others had a downward bias. This high degree of uncertainty may be disappointing to data users but we believe it is an accurate reflection of reality that must be conveyed to those who use population forecasts for decision making purposes.

For counties and other subnational areas, we believe an empirical approach is likely to provide more reliable estimates of uncertainty than models that incorporate the stochastic nature of the forecasting process. Model-based prediction intervals require a substantial amount of base data and are subject to errors in specifying the model, errors in estimating the model's parameters, and future structural changes that invalidate the model's parameter estimates (Lee, 1992). In addition, many different models can be specified, each providing a different set of prediction intervals (Cohen, 1986; Keilman, Pham, and Hetland 2002; Sanderson 1995).

Empirically-based prediction intervals have their own limitations, of course. We found that more than 90% of APEs fell inside the 90% prediction intervals in some target years and less than 90% in other target years. Intervals based on 75thPEs did not perform as well as intervals based on 90thPEs. Furthermore, the empirical approach does not provide reliable forecasts of the likely direction of future forecast errors. Further research comparing the performance of model-based and empirical prediction intervals is needed before we can draw firm conclusions regarding which approach is likely to provide more useful measures of uncertainty.

Other questions related to empirical prediction intervals remain to be answered as well. Can formal criteria be established for evaluating the stability of error distributions over time? How much historical data are needed to develop the most stable intervals? Can techniques be developed for adjusting prediction intervals to account simultaneously for the impact of differences in population size, growth rates, geographic region, and perhaps other factors as well? How do differences in the choice of cut-off points (e.g., 90th vs. 75th percentile) affect the accuracy of forecast error predictions? Can information on the distribution of errors for one geographic region be used to develop prediction intervals for another geographic region? Would results based on other forecasting techniques be similar to those reported in this study? We believe future research will provide answers to these and similar questions and enhance our ability to construct empirical prediction intervals to accompany population forecasts.

Endnotes

1. See Rayer, 2004, for a complete description of the data set.

2. We do not show coefficients of variation for algebraic percent errors because the measure loses its meaning and usefulness when the mean approaches zero and the distribution contains both positive and negative values (Lohninger, 1999).

References

Alho, J., & Spencer, B. 1990. Error Models for Official Mortality Forecasts. *Journal of the American Statistical Association*, 85, 609-616.

Armstrong, J. S. 2001. *Principles of Forecasting*. Boston: Kluwer Academic Publishers.

Cohen, J. 1986. Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-based and Empirical Approaches. *Demography*, 23, 105-126.

European Central Bank. 2006. Coefficients of Cross-County Variation. http://www.ecb.int/stats/money/interest/coeff/html/index.en.html.

Hollmann, F, Mulder, T, and Kallan, J. 2000. Methodology and Assumptions for the Population Projections of the United States: 1999 to 2100. Population Division Working Paper No. 38. U.S. Census Bureau, Washington, DC.

Hopkins, W. 2000. Measures of Reliability in Sports Medicine and Science. *Sports Medicine*, 30, 1-15.

Isserman, A. 1977. The Accuracy of Population Projections for Subcounty Areas. J*ournal of the American Institute of Planners*, 43, 247-259.

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

Keilman, N., Pham, D., and Hetland, A. 2002. Why Population Forecasts Should Be Probabilistic – Illustrated by the Case of Norway. *Demographic Research* 6, 409-453.

Keyfitz, N. 1972. On Future Population. *Journal of the American Statistical Association*, 67, 347-362.

Keyfitz, N. 1981. The Limits of Population Forecasting. *Population and Development Review*, 7, 579-593.

Lee, R. 1992. Stochastic Demographic Forecasting. *International Journal of Forecasting* 8, 315-327.

Lohninger, H. 1999. *Teach/Me Data Analysis*. Berlin: Springer-Verlag.

Lutz, W., Sanderson, W., and Scherbov, S. 1999. Expert-based Probabilistic Population Projections. In W. Lutz, J. Vaupel, & D. Ahlburg (Eds.), *Frontiers of Population Forecasting* (pp. 139-155). New York, NY: The Population Council. (A supplement to *Population and Development Review*, 24).

McKibben, J. 1996. The Impact of Policy Changes on Forecasting for School Districts. *Population Research and Policy Review* 15, 527-536.

Murdock, S., Jones, L., Hamm, S., Hwang, S., and Parpia, B. 1984. An Assessment of the Accuracy of a Regional Demographic and Economic Projection Model. *Demography*, 21, 383-404.

Pflaumer, P. 1992. Forecasting U.S. Population Totals with the Box-Jenkins Approach. *International Journal of Forecasting*, 8, 329-338.

Rayer, S. 2004. County Population Projections: An Evaluation Using One Hundred Years of Data. Paper presented at the annual meeting of the Southern Demographic Association, Hilton Head, SC.

Sanderson, W. 1995. Predictability, Complexity, and Catastrophe in a Collapsible Model of Population, Development, and Environmental Interactions. *Mathematical Population Studies* 5, 259-279.

Smith, S. 1987. Tests of Forecast Accuracy and Bias for County Population Projections. *Journal of the American Statistical Association*, 82, 991-1003.

Smith, S. and Sincich, T. 1988. Stability over Time in the Distribution of Forecast Errors. *Demography*, 25, 461-473.

Smith, S., Tayman, J., and Swanson, D. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

Sorensen, J. 2002. The Use and Misuse of the Coefficient of Variation in Organizational Demographic Research. *Sociological Methods and Research*, 30, 475-491.

Stoto, M. 1983. The Accuracy of Population Projections. *Journal of the American Statistical Association*, 78, 13-20.

Tayman, J., Schafer, E., and Carter, L. 1998. The Role of Population Size in the Determination and Prediction of Population Forecast Errors: An Evaluation using Confidence Intervals for Subcounty Areas. *Population Research and Policy Review*, 17, 1-20.

Tayman, J., Parrott, B., and Carnevale, S. 1994. Locating Fire Station Sites: The response time component. Pp. 203-217 in H. Kintner, P. Voss, P. Morrison, and T. Merrick (eds.), *Applied Demographics: A Casebook for Business and Government*. Boulder, CO: Westview Press.

Texas Water Development Board. 1997. *Water for Texas: A Consensus-based Update to the State Water Plan*. Volume 2, technical planning appendix, GF-6-2: Austin, TX.

Thomas, R. 1994. Using Demographic Analysis in Health Services Planning: A Case Study in Obstetrical Services. Pp. 159-179 in H. Kintner, P. Voss, P. Morrison, and T. Merrick (eds.), *Applied Demographics: A Casebook for Business and Government*. Boulder, CO: Westview Press.

White, H. 1954. Empirical Accuracy of the Accuracy of Selected Methods for Projecting State Populations. *Journal of the American Statistical Association*, 49, 480-498.

Williams, W. and Goodman, M. 1971. A Simple Method for the Construction of Empirical Confidence Limits for Economic Forecasts. *Journal of the American Statistical Association* 66, 752-754.

Table 1. Target Years by Launch Year and Forecast Horizon

	Forecast Horizon (Years)		
Launch Year	10	20	30
1920	1930	1940	1950
1930	1940	1950	1960
1940	1950	1960	1970
1950	1960	1970	1980
1960	1970	1980	1990
1970	1980	1990	2000
1980	1990	2000	---
1990	2000	---	---

Table 2. Forecast Errors by Target Year and Length of Forecast Horizon

Target Year	Horizon Length	MAPE	MEDAPE	MALPE	MEDALPE	90thPE	CV
1930	10	12.2	7.8	2.2	1.4	29.1	113.6
1940	10	11.2	7.9	0.4	-0.9	23.3	111.2
1950	10	11.2	7.8	2.9	2.3	24.9	99.8
1960	10	10.3	7.4	0.3	0.3	23.2	95.7
1970	10	9.6	6.8	-2.4	-2.0	21.0	101.1
1980	10	13.2	11.2	-9.5	-9.2	26.3	77.3
1990	10	7.8	6.4	4.0	4.6	15.6	85.2
2000	10	6.2	4.5	-3.5	-3.0	13.9	96.9
Average	10	10.2	7.5	-0.7	-0.8	22.2	97.6
St. Dev.	10	2.3	1.9	4.4	4.2	5.2	12.1
CV	10	22.5	25.1	-	-	23.4	12.4
1940	20	20.2	12.7	5.9	0.6	47.1	112.9
1950	20	19.9	14.3	3.6	1.0	43.6	100.4
1960	20	23.0	16.7	6.4	3.1	50.9	92.6
1970	20	16.7	12.0	-0.5	-1.2	37.3	97.5
1980	20	21.4	17.4	-12.1	-12.0	45.6	80.8
1990	20	19.4	16.1	-9.3	-9.8	39.6	77.5
2000	20	11.4	8.7	0.7	1.9	24.6	93.5
Average	20	18.9	14.0	-0.8	-2.4	41.3	93.6
St. Dev.	20	3.8	3.1	7.3	6.0	8.6	11.9
CV	20	20.1	22.1	-	-	20.9	12.8
1950	30	33.1	20.9	14.0	4.1	78.7	113.0
1960	30	32.9	23.4	8.5	2.6	68.1	104.9
1970	30	31.9	23.2	9.0	3.2	68.3	97.2
1980	30	22.1	16.8	-9.8	-10.4	49.3	90.0
1990	30	29.3	24.7	-11.9	-13.9	60.6	78.4
2000	30	27.8	24.2	-14.5	-17.4	55.3	72.8
Average	30	29.5	22.2	-0.8	-5.3	63.4	92.7
St. Dev.	30	4.2	2.9	12.6	9.7	10.5	15.4
CV	30	14.2	13.2	-	-	16.6	16.6

Table 3. Percentage of APEs Less Than the Predicted 90th Percentile Error

Target Year	Horizon Length	Percentage
1940	10	93.0
1950	10	88.4
1960	10	90.9
1970	10	92.2
1980	10	80.1
1990	10	98.5
2000	10	92.4
Average	10	90.8
St. Dev	10	5.6
CV	10	6.2
1950	20	91.5
1960	20	84.9
1970	20	95.3
1980	20	83.4
1990	20	94.0
2000	20	97.6
Average	20	91.1
St. Dev.	20	5.7
CV	20	6.3
1960	30	92.9
1970	30	89.9
1980	30	96.9
1990	30	82.0
2000	30	93.1
Average	30	91.0
St. Dev.	30	5.6
CV	30	6.2

Table 4. 90th Percentile Errors by Population Size

Target Year	Horizon Length	< 5,000	5,000 - 15,000	15,000 - 50,000	> 50,000	All
1930	10	59.9	30.1	22.8	21.9	29.1
1940	10	59.0	28.8	17.0	18.4	23.3
1950	10	45.7	26.0	22.8	22.3	24.9
1960	10	31.8	23.5	21.5	26.7	23.2
1970	10	34.8	23.4	18.7	17.4	21.0
1980	10	36.8	28.7	24.1	19.0	26.3
1990	10	19.0	16.1	14.9	15.4	15.6
2000	10	21.0	14.7	12.3	12.5	13.9
Average	10	38.5	23.9	19.3	19.2	22.2
St. Dev	10	15.5	5.8	4.2	4.4	5.2
CV	10	40.3	24.2	22.0	23.0	23.4
1940	20	86.7	57.0	35.9	35.0	47.2
1950	20	76.7	50.2	35.4	31.6	43.6
1960	20	80.8	56.1	48.1	43.2	50.9
1970	20	51.5	35.5	35.5	37.7	37.3
1980	20	66.8	52.6	38.8	32.0	45.7
1990	20	53.2	41.9	34.6	34.6	39.6
2000	20	32.3	25.9	21.8	23.8	24.6
Average	20	64.0	45.6	35.7	34.0	41.3
St. Dev.	20	19.4	11.7	7.7	6.0	8.6
CV	20	30.2	25.6	21.6	17.5	20.9
1950	30	124.7	96.5	61.9	54.6	78.7
1960	30	100.3	80.7	59.4	50.5	68.1
1970	30	115.1	73.9	64.0	55.5	68.3
1980	30	68.9	50.5	46.1	42.8	49.3
1990	30	78.1	67.4	51.7	49.3	60.6
2000	30	71.5	58.5	47.9	49.2	55.3
Average	30	93.1	71.3	55.2	50.3	63.4
St. Dev.	30	23.7	16.4	7.6	4.6	10.5
CV	30	25.4	23.0	13.7	9.1	16.6

Table 5. Percentage of APEs Less Than the Predicted 90th Percentile Error by Population Size

Target Year	Horizon Length	< 5,000	5,000 - 15,000	15,000 - 50,000	> 50,000	All
1940	10	90.5	91.4	96.3	94.1	93.0
1950	10	95.0	92.2	82.8	86.5	88.4
1960	10	96.1	91.1	91.2	87.6	90.9
1970	10	87.7	89.7	93.2	98.1	92.2
1980	10	85.8	78.2	80.3	86.5	80.1
1990	10	97.6	98.4	99.0	96.2	98.5
2000	10	86.8	91.1	93.8	94.7	92.4
Average	10	91.3	90.3	90.9	92.0	90.8
St. Dev	10	4.8	6.0	6.9	4.9	5.6
CV	10	5.3	6.7	7.6	5.4	6.2
1950	20	93.0	93.4	90.9	94.8	91.5
1960	20	88.7	87.2	80.8	81.0	84.9
1970	20	98.7	97.5	95.4	92.1	95.3
1980	20	78.4	72.2	87.2	95.0	83.4
1990	20	97.4	96.3	93.4	86.9	94.0
2000	20	97.6	97.9	97.3	97.5	97.6
Average	20	92.3	90.7	90.8	91.2	91.1
St. Dev.	20	7.8	9.9	6.0	6.2	5.7
CV	20	8.4	10.9	6.7	6.8	6.3
1960	30	88.4	94.6	91.1	93.0	92.9
1970	30	90.6	91.9	88.0	88.0	89.9
1980	30	98.0	98.4	96.9	95.0	96.9
1990	30	80.7	74.5	86.9	85.0	82.0
2000	30	94.7	95.7	91.9	90.6	93.1
Average	30	90.5	91.0	91.0	90.3	91.0
St. Dev.	30	6.6	9.5	3.9	4.0	5.6
CV	30	7.3	10.5	4.3	4.4	6.2

Table 6. 90th Percentile Errors by Population Growth Rate

Target Year	Horizon Length	< -10%	-10% to 10%	10% to 25%	> 25%	All
1930	10	35.6	17.3	27.2	44.5	29.1
1940	10	40.7	16.7	21.6	54.1	23.3
1950	10	19.6	21.6	27.6	37.8	24.9
1960	10	16.1	19.2	29.7	37.9	23.2
1970	10	23.9	17.2	17.5	31.2	21.0
1980	10	31.8	24.1	23.3	30.8	26.3
1990	10	12.4	13.5	16.1	20.6	15.6
2000	10	17.3	10.7	13.6	22.6	13.9
Average	10	24.7	17.5	22.1	34.9	22.2
St. Dev	10	10.2	4.3	5.9	11.1	5.2
CV	10	41.4	24.3	26.8	31.8	23.4
1940	20	69.7	25.4	44.8	86.7	47.2
1950	20	61.0	32.2	51.0	81.5	43.6
1960	20	44.9	45.4	57.8	74.0	50.9
1970	20	30.1	30.5	51.1	60.2	37.3
1980	20	56.8	35.2	34.7	50.5	45.7
1990	20	47.7	32.8	38.6	50.9	39.6
2000	20	27.9	19.5	25.5	37.4	24.6
Average	20	48.3	31.6	43.4	63.0	41.3
St. Dev.	20	15.5	8.1	11.1	18.2	8.6
CV	20	32.2	25.6	25.7	28.9	20.9
1950	30	71.9	43.5	82.6	145.0	78.7
1960	30	71.2	54.3	87.8	139.6	68.1
1970	30	61.6	60.6	82.6	136.6	68.3
1980	30	56.1	40.4	55.8	68.2	49.3
1990	30	72.2	46.5	53.0	70.1	60.6
2000	30	66.0	46.5	52.9	69.9	55.3
Average	30	66.5	48.6	69.1	104.9	63.4
St. Dev.	30	6.6	7.5	16.8	39.0	10.5
CV	30	9.9	15.3	24.3	37.2	16.6

Table 7. Percentage of APEs Less Than the Predicted 90th Percentile Error by Population Growth Rate

Target Year	Horizon Length	< -10%	-10% to 10%	10% to 25%	> 25%	All
1940	10	84.1	91.1	96.1	91.5	93.0
1950	10	93.8	82.9	85.0	98.6	88.4
1960	10	84.7	93.0	88.4	88.2	90.9
1970	10	75.3	90.5	95.8	93.0	92.2
1980	10	72.4	76.4	83.9	95.6	80.1
1990	10	99.5	98.6	97.9	97.0	98.5
2000	10	78.7	93.2	93.8	89.1	92.4
Average	10	84.1	89.4	91.6	93.3	90.8
St. Dev	10	9.8	7.4	5.7	4.0	5.6
CV	10	11.7	8.3	6.2	4.2	6.2
1950	20	93.7	84.7	93.6	93.1	91.5
1960	20	92.1	79.1	88.3	94.1	84.9
1970	20	91.0	96.7	95.5	91.5	95.3
1980	20	52.5	78.5	97.0	94.2	83.4
1990	20	95.7	91.7	90.1	93.2	94.0
2000	20	98.1	97.2	97.7	97.7	97.6
Average	20	87.2	88.0	93.7	94.0	91.1
St. Dev.	20	17.2	8.4	3.8	2.1	5.7
CV	20	19.7	9.6	4.1	2.2	6.3
1960	30	89.7	85.4	95.4	92.7	92.9
1970	30	92.7	86.1	95.4	93.8	89.9
1980	30	91.0	96.8	98.9	98.6	96.9
1990	30	67.8	77.2	90.3	89.5	82.0
2000	30	95.7	89.8	94.1	92.9	93.1
Average	30	87.4	87.1	94.8	93.5	91.0
St. Dev.	30	11.2	7.1	3.1	3.3	5.6
CV	30	12.8	8.2	3.3	3.5	6.2

Appendix 1: Trend Extrapolation Techniques

We used the following forecasting techniques: linear (LIN), modified linear (MLN), share-of-growth (SHR), shift-share (SFT), exponential (EXP), constant-share (COS), and constant (CON). The linear technique (LIN) assumes that the population will increase (decrease) by the same number of persons in each future decade as the average per decade increase (decrease) observed during the base period:

1) $P_t = P_l + x / y (P_l – P_b)$,

where P_t is the population in the target year, P_l is the population in the launch year, P_b is the population in the base year, x is the number of years in the forecast horizon, and y is the number of years in the base period.

The modified linear technique (MLN) initially equals the linear technique, but in addition distributes the difference between the sum of the linear county forecasts and an independent national forecast proportionally by population size at the launch year:

2) $P_{it} = LIN + P_{il} / P_{jl} (P_{jt} – \Sigma LIN)$,

where i represents the county and j the nation.

The share-of-growth technique (SHR) assumes that each county's share of population growth will be the same over the forecast horizon as it was during the base period:

3) $P_{it} = P_{il} + [(P_{il} – P_{ib}) / (P_{jl} – P_{jb})] (P_{jt} – P_{jl})$,

while the shift-share technique (SFT) assumes that the average per decade change in each county's share of the national population observed during the base period will continue throughout the forecast horizon:

4) $P_{it} = P_{jt} [P_{il} / P_{jl} + (x / y) (P_{il} / P_{jl} – P_{ib} / P_{jb})]$.

The exponential technique (EXP) assumes the population will grow (decline) by the same rate in each future decade as during the base period:

5) $P_t = P_l e^{rx}$, $r = [\ln (P_l / P_b)] / y$,

where e is the base of the natural logarithm and ln is the natural logarithm.

The constant-share technique (COS) assumes the county's share of the national population will be the same in the target year as it was in the launch year:

6) $P_{it} = (P_{il} / P_{jl}) P_{jt}$,

while the constant technique (CON) assumes that the county population in the target year is the same as in the launch year:

7) $P_t = P_l$

Four of these techniques (MLN, SFT, SHR, and COS) require an independent national forecast for the target year population. Since no set of national forecasts covers all the launch years and forecast horizons used in this study, we constructed a set by applying the linear and exponential techniques to the national population. We used an average of these two forecasts as a forecast of the U.S. population.