Demography, Vol. 28, No. 2, May 1991

An Empirical Analysis of the Effect of Length of Forecast Horizon on Population Forecast Errors

Stanley K. Smith
Department of Economics and
Bureau of Economic and Business Research
University of Florida
Gainesville, FL 32611

Terry Sincich
Department of Information Systems
and Decision Sciences
University of South Florida
Tampa, FL 33620

Many studies have found that population forecast errors generally increase with the length of the forecast horizon, but none have examined this relationship in detail. Do errors grow linearly, exponentially, or in some other manner as the forecast horizon becomes longer? Does the error-horizon relationship differ by forecasting technique, launch year, size of place, or rate of growth? Do alternative measures of error make a difference? In this article we address these questions using two simple forecasting techniques and population data from 1900 to 1980 for states in the United States. We find that in most instances there is a linear or nearly linear relationship between forecast accuracy and the length of the forecast horizon, but no consistent relationship between bias and the length of the horizon. We believe that these results provide useful information regarding the nature of population forecast errors.

Many empirical studies have found population forecast errors to increase with the length of the forecast horizon (e.g., Kale, Voss, Palit, and Krebs 1981; Keyfitz 1981; Schmitt and Crosetti 1951; Smith 1987; White 1954). To our knowledge, however, no study has examined this relationship in detail. Do errors grow linearly, exponentially, or in some other manner as the forecast horizon becomes longer? Is the relationship between error and length of horizon the same for all forecasting techniques, time periods, and measures of forecast error? Do population size and rate of growth make any difference? In the present study we provide some preliminary answers to these questions.

These questions are important because decisions involving billions of dollars are based (at least in part) on population forecasts. Planning for schools, hospitals, shopping centers, roads, housing developments, and many other projects is dependent upon expected population changes. Yet decision makers often have little idea of the potential accuracy of the population forecasts they are using. The study of past forecast errors can provide some indication of potential future errors and can show how those errors may be expected to change as the length of the forecast horizon increases. We believe this information will be useful for many types of decision making.

A population forecast is defined in this study as the future population value produced by a particular forecasting technique and set of base data. The following terms are used to describe population forecasts:

Base year: The year of the earliest observed population size used to make a forecast;
Launch year: The year of the latest observed population size used to make a forecast;
Target year: The year for which population size is forecast;
Base period: The interval between the base year and the launch year;
Forecast horizon: The interval between the launch year and the target year.

Data and Techniques

The data used in this study were taken from Census Bureau reports showing decennial census counts and annual intercensal estimates for states in the United States from 1900 to 1980 (U.S. Bureau of the Census 1956, 1965, 1971, 1976, 1982, 1984). These reports covered all states (including the District of Columbia) from 1950 onward, and all states except Alaska and Hawaii from 1900 to 1949. The data refer to the total resident population on July 1 of each year.

The intercensal estimates made by the Census Bureau were based on statistical series that reflect changes in population size. For all decades, annual data on births, deaths, and school enrollment were used. For some decades, five-year migration data from the decennial census were used as well. In a few instances, data from special censuses were included. In recent years, data from federal income tax returns and Medicare records have been included. All intercensal estimates were controlled to ensure that they were consistent with decennial census counts. Although these estimates certainly contain some errors (especially for years before 1930), we believe they are quite reliable and provide a sound basis for producing population forecasts.

We used four simple extrapolation techniques to produce population forecasts; for purposes of brevity, only two are discussed in this article.[1] The first was linear extrapolation (LINE), which assumes that a population will increase (decrease) by the same number of persons in each future year as the average annual increase (decrease) during the base period:

$$\hat{P}_t = P_o + x/y\ (P_o - P_b) \qquad (1)$$

where $\hat{P}_t$ = state population forecast for the target year, P_o = state population in the launch year, P_b = state population in the base year, x = number of years in the forecast horizon, and y = number of years in the base period.

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

$$\hat{P}_t = P_o \exp(rx) \qquad (2)$$

where r = average annual growth rate during the base period.

Simple techniques such as these are used frequently for small-area population forecasts but are no longer common for state forecasts, having been replaced by more sophisticated cohort-component and economic-demographic techniques. These more sophisticated techniques, however, have been used only within the last several decades. Thus no state forecasts employing these techniques have been produced for most of the decades of this century. Creating a set specifically for this study would have been difficult or even impossible, given the lack of relevant historical data. The main advantage of the simple techniques is that they can be used to create consistent sets of forecasts for all states for many different launch years and forecast horizons, as required by the present study. Although they are not useful for some purposes, we believe that simple extrapolation techniques provide a useful basis for investigating the relationship between population forecast errors and the length of the forecast horizon.

In addition, a number of studies have found population forecast errors from simple techniques to be very similar to those from more sophisticated techniques (e.g., Ascher 1981; Kale et al. 1981; Siegel 1953; Smith 1984; White 1954). Thus the findings reported in this study may apply to other techniques as well. Further research must be performed before we can draw general conclusions, but the present study points to some of the directions that research might take.

Empirical Analysis

Using these techniques and population data from 1900 to 1980, we made forecasts with horizons expanding in five-year intervals from five to 50 years. The total number of forecasts was limited by using as launch years only years since 1910 ending in 0 or 5. We used a base period of 10 years for all forecasts; a previous study showed that approximately 10 years of base data are necessary (and generally sufficient) to achieve the highest possible degree of accuracy for these forecasting techniques (Smith and Sincich 1990). We replicated the forecasts using 20-year base periods; the results were very similar to those reported in this article.

Forecast error (F_t) is defined as the percentage difference between the population forecast ($\hat{P}_t$) and the "true" population (P_t) in the target year:

$$F_t = \left(\frac{\hat{P}_t - P_t}{P_t}\right) 100. \tag{3}$$

We assumed that "true" population numbers were those published by the U.S. Bureau of the Census; that is, we made no attempt to adjust the data for estimation or enumeration error.

We evaluated six measures of forecast accuracy and bias; only two are reported in this article.[2] Mean absolute percentage error (MAPE) is the average error when the direction of the error is ignored. This provides a measure of accuracy. Mean algebraic percentage error (MALPE) is the average percentage error when the direction of the error is accounted for. This provides a measure of bias: a positive error indicates that forecasts tend to be too high and a negative error indicates that forecasts tend to be too low.

Results from Aggregate Data

For our first analysis we aggregated forecast errors across all launch years. Table 1 shows the number of states for which forecasts were made for each horizon, both for all states and for states divided according to the population size in the launch year and the population growth rate during the base period. Figure 1 summarizes the empirical results.

For the LINE technique, the MAPE grew in approximately linear fashion as the horizon increased to 35 years, and then dipped slightly before increasing further. For MALPE the pattern was completely different. Values were negative for all horizons; they declined as the horizon increased to 35 years and then moved back toward zero. For the EXPO technique, the MAPE grew approximately linearly as the horizon increased to 35 years, but then it began to increase at an increasing rate. The MALPE was positive for all forecast horizons and grew at an increasing rate after the horizon passed 30 years; this

Table 1. Number of State Forecasts, by Length of Horizon, Population Size, and Growth Rate

		Growth Rate ≥ 25%		Growth Rate < 25%	
Forecast Horizon	Total	<1 million	≥1 million	<1 million	≥1 million
5	694	66	71	153	404
10	643	63	68	143	367
15	592	61	63	131	337
20	541	57	57	129	397
25	490	51	50	110	279
30	441	44	37	192	258
35	392	49	32	91	229
40	343	37	28	79	199
45	294	36	27	65	166
50	245	35	25	51	134

pattern reflects a strong tendency for EXPO forecasts to be too high, especially for longer forecast horizons. This finding will be explained later in this article.

Do the patterns shown in Figure 1 hold in general, or only for certain time periods or for states with particular population characteristics? To answer this question, we must consider forecast errors for individual launch years and for states in various size/growth categories.

Results by Launch Year

Figure 2 shows MAPEs by forecast horizon for launch years ending in 0. We omitted launch years ending in 5 to avoid clutter in the diagrams. The results for all omitted years except 1945 were similar to those shown here; for 1945, errors were considerably larger than for other launch years.

For the LINE technique, errors increased approximately linearly with the forecast horizon for all seven sets of forecasts. The only exception to this pattern was an upward deviation from the linear trend found in all forecasts for target year 1945; World War II apparently had a major impact on the accuracy of population forecasts for the 1940s. Figure 2 gives no indication that errors for LINE tend to level off after the horizon reaches 35 years. The leveling off shown in Figure 1 was most likely the result of very large errors for forecasts for 1945 and of the impact of those errors on the averages for horizons of 35 years and less (no horizons of longer than 35 years had 1945 as a target year).

For the EXPO technique, we found an approximately linear relationship between MAPE and the forecast horizon for all launch years after 1910, except for the upward deviation found again for 1945. For launch year 1910, however, the relationship between MAPE and length of horizon was clearly nonlinear. (Some of the values were too large to fit into Figure 2; for example, the MAPE reached 250% for the 50-year forecast.) The results for 1910 will be explained later.

The analysis of results for individual launch years implies that the increasing slope of the MAPE-horizon line for EXPO (shown in Figure 1) may have been caused by the very large errors in the forecasts with 1910 (and, to a lesser extent, 1915) as a launch year. To test for this possibility, we omitted forecasts based on these two launch years and

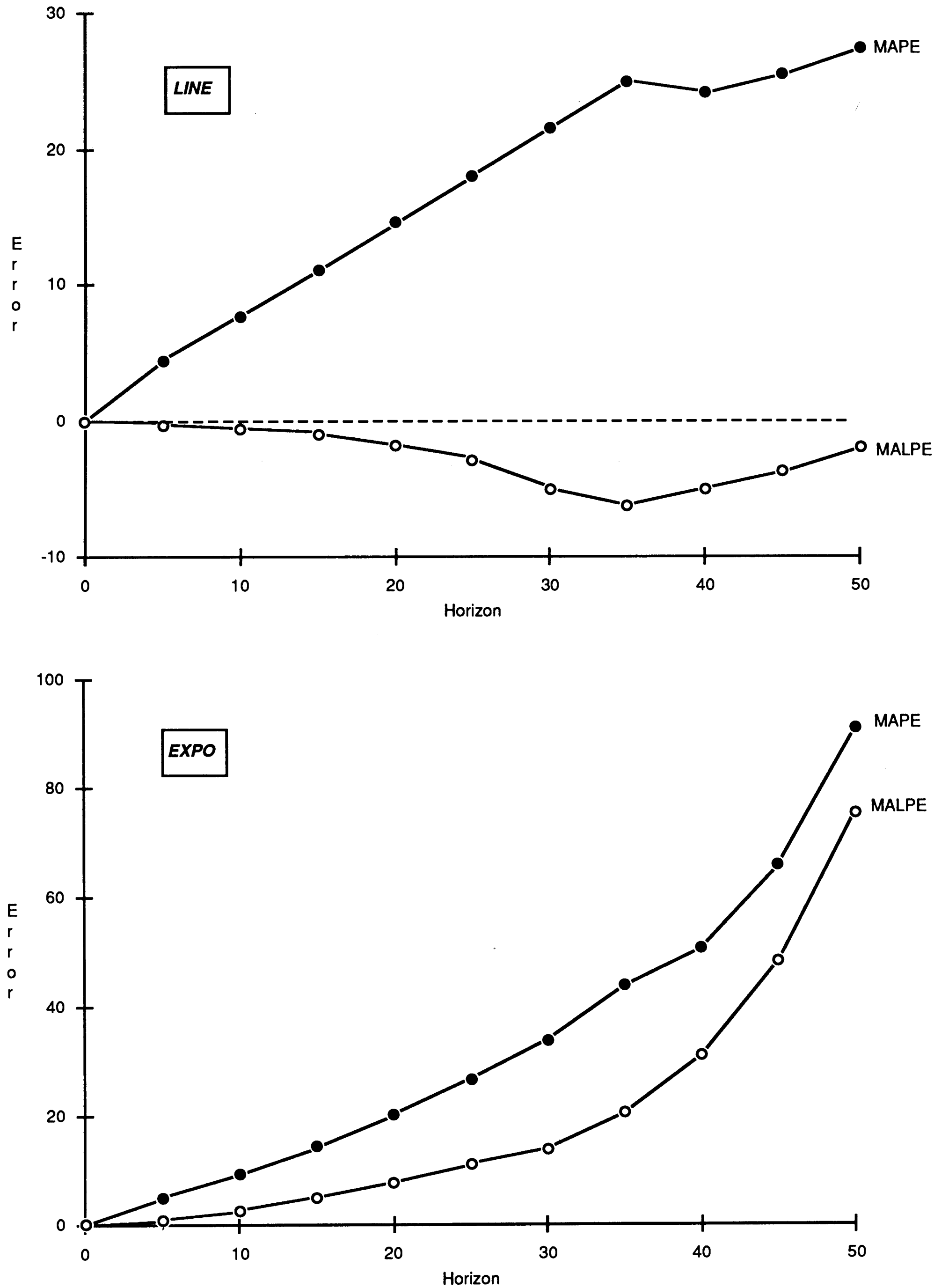

Figure 1. Forecast Error, by Length of Horizon

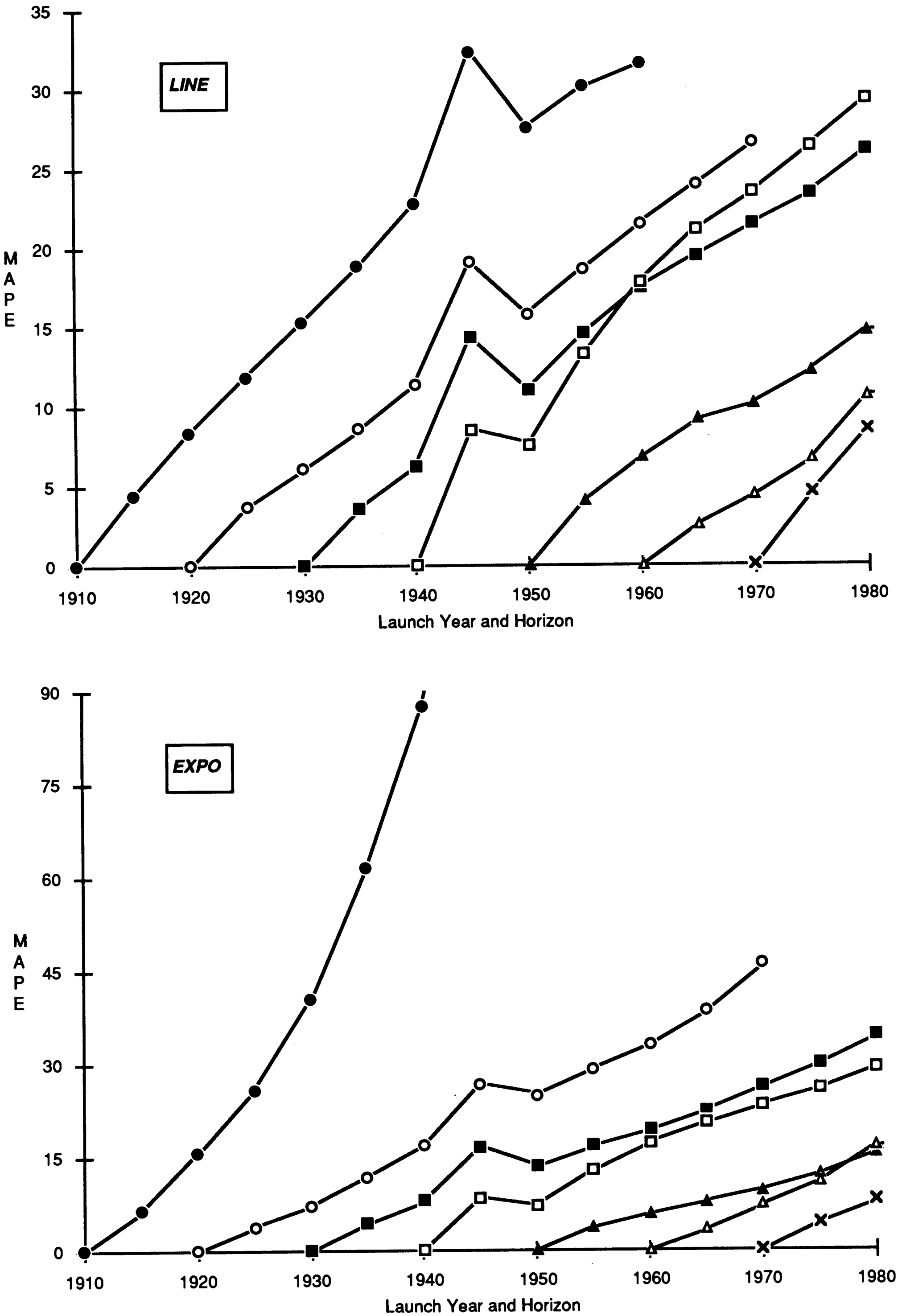

Figure 2. MAPE, by Launch Year and Forecast Horizon

recalculated the errors summarized in Figure 1. We found that the MAPE-horizon relationship for EXPO was now approximately linear over the entire 50-year horizon.

We also evaluated MALPEs by individual launch year (not shown here), but we found no consistent patterns. For some launch years, MALPEs were positive for all forecast horizons; for others, they were negative for all horizons; and for still others, they were positive for some horizons and negative for other horizons. In some instances MALPEs increased with the forecast horizon; in others, they declined. As noted before, there appear to be no general biases in these population forecasting techniques (Smith and Sincich 1988).

Results by Population Size and Growth Rate

A number of studies have found forecast errors to vary by population size and/or growth rate (e.g., Keyfitz 1981; Schmitt and Crosetti 1951; Smith 1987; White 1954). To investigate these effects, we divided states according to population size in the launch year (< 1 million, ≥ 1 million) and growth rate during the base period (< 25%, ≥ 25%), producing four size-growth categories. The number of state forecasts in each category is shown in the last four columns of Table 1.

Figure 3 shows MAPEs by size-growth category. For LINE, errors were consistently largest for small, rapidly growing states and smallest for large, slowly growing states. In all four categories, however, errors grew in approximately linear fashion with increases in the forecast horizon to 35 years, and then leveled off somewhat. For LINE, then, the results for each category were similar to those reported in Figure 1 for the entire sample of states.

For EXPO, errors also were largest for rapidly growing states (especially small states) and smallest for slowly growing states (especially large states). The MAPE-horizon relationship was basically linear for slowly growing states (dipping slightly after 35 years), but was more nearly exponential for rapidly growing states. We believe that forecast errors for EXPO in rapidly growing states increased at an increasing rate as the forecast horizon became longer because high growth rates tend to regress toward the mean over time, whereas the EXPO technique forecasts those rates to remain constant (Smith 1987).

The results shown in Figure 3 provide insight into some of the nonlinear relationships for EXPO shown in Figures 1 and 2. The increasing slopes of the trend lines observed in Figure 1 were not characteristic of EXPO forecasts for all states; rather, they reflected the large errors for rapidly growing states. In addition, the large errors for the EXPO forecasts for launch year 1910 (Figure 2) most likely were caused by the large number of small states that existed in 1910 and the large number of states that grew very rapidly between 1900 and 1910: 20 states had populations of less than one million in 1910 and 20 states had growth rates greater than 25% between 1900 and 1910 (11 states had growth rates greater than 50%).

Statistical Analysis of MAPE-Horizon Relationship

Figures 1, 2, and 3 provide visual evidence that except for EXPO forecasts of rapidly growing states, the relationship between the MAPE and the length of the forecast horizon in this sample is approximately linear. These figures, however, do not account for variation around the means. Consequently we cannot measure the reliability of inferences derived from these figures. To produce such measures, formal statistical tests are required. One approach is to use multiple regression analysis to model the MAPE-horizon relationship. We propose three regression models:

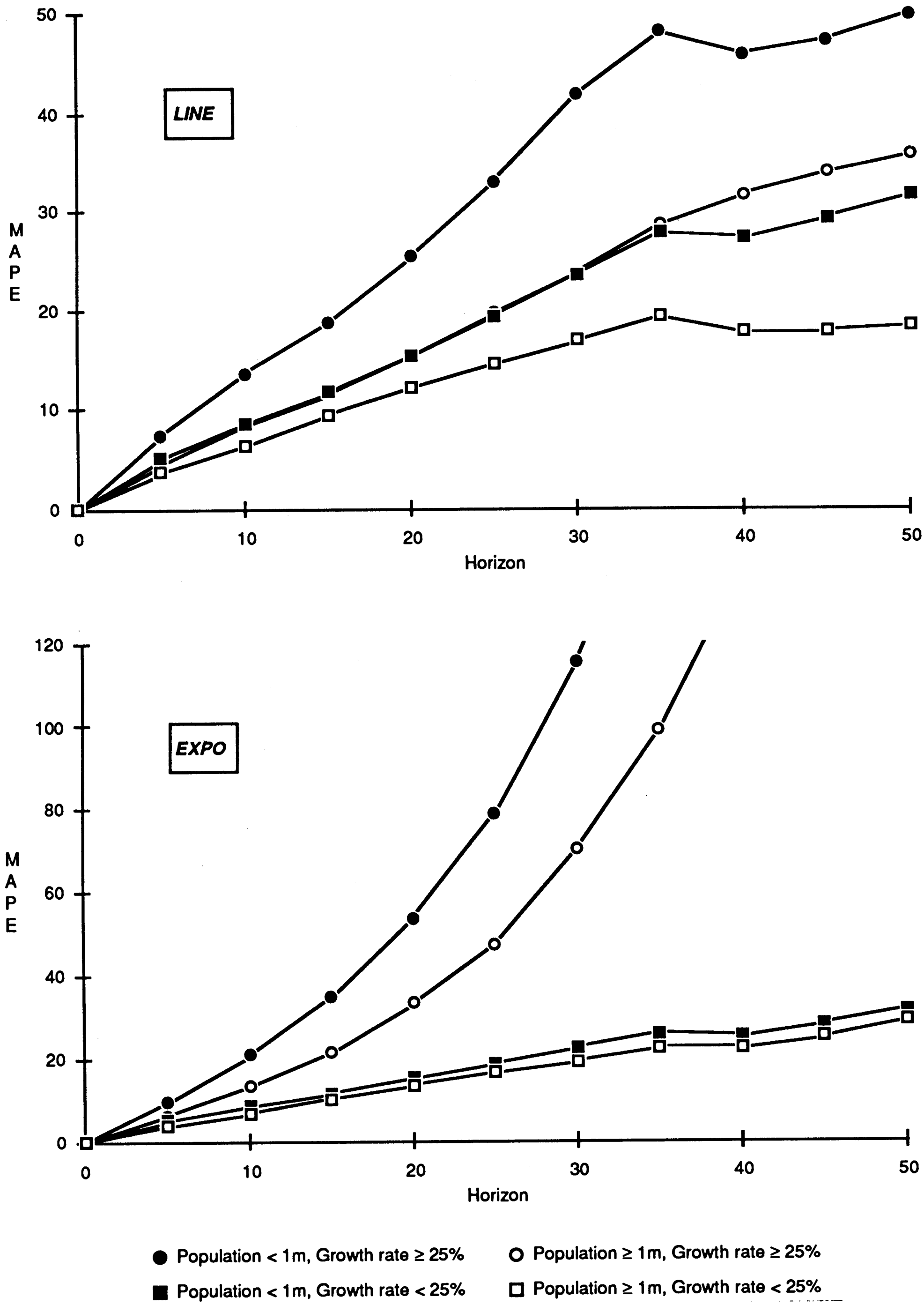

Figure 3. MAPE, by Length of Horizon for Size-Growth Categories

$$\text{Model L (linear): MAPE} = \beta_0 + \beta_1 x + \epsilon \quad (5)$$

$$\text{Model Q (quadratic): MAPE} = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \quad (6)$$

$$\text{Model E (exponential): MAPE} = \text{exponential}\ (\beta_0 + \beta_1 x + \epsilon) \quad (7)$$

where x = number of years in the forecast horizon and ϵ = random error (assumed to be distributed normally with mean 0 and variance ς^2).[3]

A direct comparison of the linear and the quadratic models can be made by conducting a t-test on the curvature term in Model Q (i.e., testing the null hypothesis that $\beta_2 = 0$). A more general method is required to compare the linear and the exponential models. One widely used technique (Box and Cox 1962) leads to standard statistical inferences about the choice of the transformation of the dependent variable. For the special case of comparing Model L to Model E, this technique reduces (after making the proper normalizing power transformation of MAPE) to choosing the model with the smallest residual sum of squares. Other tests, such as the Score test (Cox and Hinkley 1974) and a test suggested by Bera and McAleer (1982), are based on artificial regressions with "constructed" variables. For each model we test the null hypothesis that the β parameter associated with the "constructed" or added variable is 0. Let θ_L and θ_E be the added variable parameters for the linear and the exponential models respectively. If we fail to reject the hypothesis $\theta_L = 0$ but reject the hypothesis $\theta_E = 0$, we choose Model L. Likewise, if we reject $\theta_L = 0$ but fail to reject $\theta_E = 0$, we choose Model E. The tests are inconclusive if both hypotheses are accepted or if both are rejected. Maddala (1988) and Atkinson (1985) provide details on how to construct the added variables for these and similar tests.

For each forecasting technique, we calculated MAPEs from the aggregate data for each of the 10 forecast horizons (5, 10 . . . 50 years). We then fitted the three regression models to these 10 data points. (Data for target year 1945 and launch year 1945 were omitted from the analysis for reasons given earlier.) Table 2 provides a summary of the results. Under Model E, the first R^2 reported is based on the difference between ln(MAPE) and the predicted value of ln(MAPE). Because the dangers of comparing this R^2 to those for Models L and Q are well known (Maddala 1988), for Model E we report a second value of R^2 (in parentheses) based on the difference between MAPE and the predicted value of MAPE, where the predicted MAPE is obtained by calculating the exponential of the predicted value of ln(MAPE). This "pseudo" R^2 value is preferred for comparing models.

For the LINE technique, Models L and Q performed considerably better than Model E with respect to adjusted R^2 values and the accuracy of predicted values. The Box-Cox residual sum of squares shows that Model L is superior to Model E; the Bera-McAleer (BM) and the Score tests also provide some support for Model L. For the EXPO technique, Models Q and E performed better than Model L with respect to the adjusted R^2 and the accuracy of the predicted values, and the Box-Cox residual sum of squares supports the choice of Model E over Model L. The BM and the Score tests, however, yield contradictory results. The BM test is not significant at the 10% level for either Model L or Model E, whereas the Score test is significant at the 1% level for both models.

For both techniques, the curvature coefficient in Model Q was found to be significantly different from 0. Thus it appears that Model Q provides the best fit of forecast errors for both forecasting techniques. For the LINE technique, however, the differences between Models Q and L were extremely small, an indication that the MAPE-horizon relationship was nearly linear. Only for EXPO did Model L clearly provide a poorer fit of the data than Model Q.

Figure 1 showed that curvilinear trends for both LINE and EXPO were due mostly to forecast errors at horizons beyond 35 years; for horizons of 35 years or less, the trends appeared to be about linear. We tested this observation by dropping horizons of 40, 45, and

Table 2. Actual and Predicted Values of MAPE for Three Regression Models, by Technique and Length of Forecast Horizon

Horizon	Actual MAPE	Predicted MAPE Model L	Model Q	Model E
LINE Technique				
5	3.5	4.3	3.2	5.5
10	6.5	7.0	6.6	6.7
15	9.4	9.7	9.9	8.3
20	12.8	12.4	12.9	10.2
25	16.1	15.1	15.8	12.6
30	18.6	17.8	18.5	15.5
35	20.7	20.5	21.0	19.1
40	24.0	23.2	23.4	23.5
45	25.4	25.9	25.5	29.0
50	27.3	28.6	27.5	35.7
R-squared (adjusted):		.990	.998	.871 (.780)
Test for curvature (p-value):			.001	
Box-Cox residual SS:		.028		.465
BM test (p-value):		.013		.001
Score test (p-value):		.060		.000
EXPO Technique				
5	4.0	−4.6	8.2	6.0
10	8.3	3.9	8.2	8.1
15	13.1	12.4	10.3	11.0
20	19.4	21.0	14.5	15.0
25	26.0	29.5	20.9	20.4
30	30.1	38.0	29.5	27.8
35	29.5	46.6	40.1	37.8
40	50.6	55.1	52.9	51.4
45	65.9	63.6	67.9	70.0
50	90.8	72.1	85.0	95.2
R-squared (adjusted):		.862	.956	.947 (.972)
Test for curvature (p-value):			.004	
Box-Cox residual SS:		1.479		.390
BM test (p-value):		.102		.263
Score test (p-value):		.000		.000

Note: Forecasts with 1945 as target year or launch year have been omitted from the analysis.

50 years from the analysis and fitting the regression models to the data for horizons of 5 to 35 years (not shown here). For both LINE and EXPO, we found that the curvature term in Model Q was not significantly different from 0 at a 5% level. Furthermore, we found that the Box-Cox residual sum of squares for Model E exceeded the values for Model L; that both the BM and the Score tests for Model L were not statistically significant at a 10% level; and that both the BM and the Score tests for Model E were statistically significant at a 1% level. These results provide strong statistical support for the observation that forecast errors grew approximately linearly as the forecast horizon increased to 35 years, even for the EXPO technique.

We also ran the regressions for MAPEs calculated by individual launch year (not shown here). For Model Q, the curvature term was found to be statistically insignificant (at 5%) for every launch year for the LINE forecasts and for every launch year but one for the EXPO forecasts. Models L and Q each performed considerably better than Model E for the LINE technique, regardless of launch year. For the EXPO technique, only MAPEs for launch year 1910 appeared to grow in a nonlinear manner, and the nonlinear trend was fit better by a quadratic function (Model Q) than by an exponential function (Model E). The results for individual launch years thus support the observation that forecast errors generally grow approximately linearly with the forecast horizon.

We replicated Table 2 with states divided into the four size-growth categories described earlier (not shown here). As with the data for all states together, Models L and Q had very similar results in all size-growth categories for the LINE technique; each performed considerably better than Model E. For the EXPO technique, Models L and Q performed better than Model E for slowly growing states (i.e., with growth rates of less than 25%). The curvature term in Model Q was small and insignificant, implying linear growth in MAPEs. For rapidly growing states (i.e., with growth rates of 25% or more), the curvature term in Model Q was statistically significant, implying nonlinear growth. Models Q and E each performed better than Model L in these states.

When we excluded forecasts with horizons of 40, 45, and 50 years from the analysis, the linear trend was supported even more strongly. The curvature term in Model Q was statistically insignificant (at 5%) in all four size-growth categories and for both forecasting techniques. Models L and Q each performed considerably better than Model E in all categories, even for the EXPO forecasting technique.

Thus the statistical analysis supports strongly our observation that the MAPE grows approximately linearly with the forecast horizon. For horizons out to 35 years, we found a linear relationship for both forecasting techniques and for all size-growth categories. For horizons out to 50 years, we found a linear relationship for all LINE forecasts and for EXPO forecasts in states with growth rates of less than 25%. Only for EXPO forecasts of rapidly growing states was it necessary to reject the hypothesis that the MAPE increases linearly with the forecast horizon.[4]

Comparison with other Studies

How do these results compare with those found in other studies? A number of studies have found forecast errors to increase approximately linearly with the length of the forecast horizon. A study of autoregressive integrated moving average (ARIMA) forecasts for states found MAPEs of 3.7% for five-year forecasts, 6.8% for 10-year forecasts, 9.0% for 15-year forecasts and 12.4% for 20-year forecasts (Kale et al. 1981). A study of 20 city population forecasts using a ratio method reported errors of 9.3% for 10-year forecasts and 15.5% for 20-year forecasts (Schmitt and Crosetti 1951). A study of 2,971 county forecasts using the LINE, EXPO, and Shift-Share techniques found MAPEs of 12 to 15% for 10-year forecasts and 25 to 35% for 20-year forecasts (Smith 1987). A study of state forecast errors using an average of several techniques found errors of 7% for 10-year forecasts and 15% percent for 20-year forecasts (White 1954). Ascher (1981) concluded that short-term forecasts were more accurate than longer forecasts, and that errors often increased in a nearly linear manner.

Keyfitz (1981) and Stoto (1983) analyzed a large number of forecasts made for countries. Instead of using the MAPE as a measure of error, they focused on the difference between the forecast rate of population increase and the actual rate realized over time. They concluded that this difference tends to remain constant over the entire length of the forecast horizon. It can be demonstrated that this finding is virtually the same as our conclusion that the MAPE grows linearly as the forecast horizon increases.[5] Thus a linear or approximately linear relationship between mean forecast error and the length of the forecast horizon has been found in studies using several different forecasting techniques, time periods, and geographic units.

Conclusions

With only a few exceptions, this study found a linear or nearly linear relationship between forecast accuracy and the length of the forecast horizon for all forecasting techniques, for all launch years, and for states in all population size-growth rate categories.[6] The only exceptions were forecasts with 1945 as the launch or target year and EXPO forecasts for rapidly growing states. (Even for the latter forecasts, the relationship was basically linear until the horizon reached 35 years.) These exceptions can be explained as follows: First, World War II caused disruptions that strongly affected population growth trends during the 1940s. Catastrophes of such magnitude and scope will cause any extrapolatory forecasting technique to err substantially. Second, EXPO forecasts are based on the assumption of constant growth rates, but very high growth rates tend to decline over time (Smith 1987). Consequently, EXPO forecasts for rapidly growing states produced errors that grew at an increasing rate over time, especially at the longer horizons. Other than these two exceptions, the present study found a very strong tendency for population forecast errors to grow approximately linearly with the forecast horizon.

This conclusion refers only to measures of forecast accuracy, not to measures of bias. MALPEs differed from one forecasting technique to another, from one size-growth category to another, from one launch year to another, and over the length of the forecast horizon. Although some regularities in the data were evident (e.g., MALPEs generally were positive for states that grew rapidly during the base period and negative for states that grew slowly), we can make no generalizations regarding bias that fit all techniques, size-growth categories, and launch years.

These conclusions—although supported strongly in the present analysis—must be accepted as preliminary because they are based on simple extrapolation techniques and a limited number of demographic contexts. Would a linear error-horizon relationship be found for other commonly used forecasting techniques, such as cohort-component, economic-demographic, and time series? Would similar results be found for cities and counties, which typically exhibit more volatility in growth rates over time than states and have a broader distribution of population forecast errors? Is the error-horizon relationship linear in the very short run and in the very long run, or only for the horizons covered in this study? Would different measures of accuracy lead to different results? What is the theoretical basis for expecting a linear error-horizon relationship? There remain many gaps in our knowledge of the relationship between population forecast errors and the length of the forecast horizon.

Perhaps the most critical question is whether a linear error-horizon relationship would be found for forecasts derived from other commonly used forecasting techniques. Although we cannot answer this question conclusively, several factors lead us to believe that other techniques generally would produce results similar to those reported here. First, although the LINE and the EXPO techniques differ greatly from each other in their assumptions regarding population growth, we found a linear relationship in most instances for both techniques.[7] Second, linear or nearly linear relationships have been reported in studies using other forecasting techniques; for example, most of the forecasts analyzed by Keyfitz (1981) and Stoto (1983) were based on cohort-component models. Third, several studies comparing simple with more sophisticated forecasting techniques have found that errors were similar for both types of techniques (e.g., Ascher 1981; Kale et al. 1981; Siegel 1953; Smith 1984; White 1954). Further research is needed before we can draw general conclusions, but we believe that the error-horizon relationship for most other forecasting techniques will be much the same as that reported in this study.

If the results reported in this article can be generalized, what will they mean for the producers and users of population forecasts? The linearity or near linearity of the error-horizon relationship means that a relatively limited amount of information on past forecast errors can be used to develop forecasts of future forecast errors. For example, suppose that data on past forecast errors for a particular technique and level of geography exist only for five- and 10-year forecasts. We can use those data as predictors of future forecast errors, not only for five- and 10-year horizons but also for 15-, 20-, and 25-year horizons. This approach will provide valuable information on the expected reliability of current population forecasts and will give decision makers an additional tool to use in planning.[8]

Notes

[1] The other two techniques were share-of-growth (whereby states are forecast to have the same share of national growth in the future as during the base period) and shift-share (whereby state shares of national population are forecast to change by the same annual amount in the future as during the base period). The empirical results for these two techniques were very similar to those reported here for LINE and EXPO.

[2] The other four measures were root mean squared percentage error and the 50th, 70th, and 90th percentile errors (i.e., the absolute percentage errors larger than exactly 50%, 70%, and 90% of all absolute percentage errors). These are all measures of forecast accuracy. In most instances the error-horizon relationship for these measures was similar to that reported here for MAPE.

[3] We included intercept terms because we wanted the models that fit the data best; those with an intercept term provided the best fit. Furthermore, regression through the origin presents problems in situations where the form of the model is unknown (Mendenhall and Sincich 1989).

[4] The assumptions regarding ϵ must be satisfied before the statistical inferences described above can be accepted as valid. With such small samples (10 or fewer data points), formal statistical tests of the assumptions would not be very powerful. In nearly all cases, however, a graphical examination of the regression residuals provided no glaring violations of either the normality or the constant-variance assumption. In addition, the sample first-order residual autocorrelations rarely exceeded .4 and generally were smaller than .2; thus the assumption of independence also appears to be satisfied to a reasonable degree.

[5] A mathematical proof is available from the authors on request.

[6] This conclusion, of course, is based on aggregate measures of error. For forecasts of any individual state, the results may well be different.

[7] In addition, we found a linear error-horizon relationship in most instances for forecasts derived from the shift-share and share-of-growth techniques.

[8] Another application of the empirical approach described in this article is the production of confidence limits for population forecasts (e.g., Keyfitz 1981; Smith 1987; Smith and Sincich 1988; Stoto 1983). Confidence limits, however, also can be derived from other approaches to population forecasting (e.g., Cohen 1986; Pflaumer 1988). Different approaches to population forecasting and to the production of confidence limits may lead to substantially different conclusions regarding the appropriate size of those limits. Future research must consider the theoretical and empirical validity of these alternative approaches.

Acknowledgments

An earlier version of this paper was presented at the annual meeting of the Population Association of America, March 30 - April 1, 1989, in Baltimore. The authors thank Ravi Bayya for valuable research assistance and Pam Middleton for skillful typing.

References

Ascher, W. 1981. "The Forecasting Potential of Complex Models." *Policy Sciences* 13: 247–67.

Atkinson, A.C. 1985. *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*. Oxford: Clarendon.

Bera, A.K. and M. McAleer 1982. "Further Results on Testing Linear and Log-Linear Regression Models." Presented at the SSRC Econometric Group Conference on Model Specification and Testing, Warwick, UK.

Box, G.E.P. and D.R. Cox 1962. "An Analysis of Transformations." *Journal of the Royal Statistical Society, Series B*: 211–43.

Cohen, J. 1986. "Population Forecasts and Confidence Intervals for Sweden: A Comparison of Model-Based and Empirical Approaches." *Demography* 23: 105–26.

Cox, D.R. and D. V. Hinkley 1974. *Theoretical Statistics*. London: Chapman and Hall.

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

Keyfitz, N. 1981. "The Limits of Population Forecasting." *Population and Development Review* 7: 579–93.

Maddala, G.S. 1988. *Introduction to Econometrics*. New York: Macmillan.

Mendenhall, W. and T. Sincich 1989. *A Second Course in Business Statistics: Regression Analysis*. London: Macmillan.

Pflaumer, P. 1988. "Confidence Intervals for Population Projections Based on Monte Carlo Methods." *International Journal of Forecasting* 4: 135- 42.

Schmitt, R. and A. Crosetti 1951. "Accuracy of the Ratio Method for Forecasting City Population." *Land Economics* 27: 346–48.

Siegel, J. S. 1953. "Forecasting the Population of Small Areas." *Land Economics* 29: 72–88.

Smith, S. K. 1984. *Population Projections: What Do We Really Know?* Gainesville: Bureau of Economic and Business Research, University of Florida.

______. 1987. "Tests of Forecast Accuracy and Bias for County Population Projections." *Journal of the American Statistical Association* 82: 991–1003.

Smith, S. K. and T. Sincich 1988. "Stability over Time in the Distribution of Population Forecast Errors." *Demography* 25: 461–74.

______. 1990. "On the Relationship between Length of Base Period and Population Forecast Errors." *Journal of the American Statistical Association* 85: 367–75.

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

U.S. Bureau of the Census 1956. "Estimates of the Population of States: 1900 to 1949." *Current Population Reports* Series P-25, 139. Washington, DC: U.S. Government Printing Office.

______. 1965. "Revised Estimates of the Population of States and Components of Change: 1950 to 1960." *Current Population Reports* Series P-25, 304. Washington, DC: U.S. Government Printing Office.

______. 1971. "Preliminary Intercensal Estimates of States and Components of Population Change: 1960 to 1970." *Current Population Reports* Series P-25, 460. Washington, DC: U.S. Government Printing Office.

______. 1976. "Estimates of the Population of States with Components of Change: 1970 to 1975." *Current Population Reports* Series P-25, 640. Washington, DC: U.S. Government Printing Office.

______. 1982. "Provisional Estimates of the Population of States: July 1, 1981." *Current Population Reports* Series P-25, 911. Washington, DC: U.S. Government Printing Office.

______. 1984. "Estimates of the Population of States: 1970 to 1983." *Current Population Reports* Series P-25, 957. Washington, DC: U.S. Government Printing Office.

White, H. R. 1954. "Empirical Study of the Accuracy of Selected Methods of Projecting State Population." *Journal of the American Statistical Association* 49: 480–98.