An Examination of the Determinants of Population Forecast Errors for Counties

Jeff Tayman, UC San Diego

Stanley K. Smith, University of Florida

Stefan Rayer, University of Florida

Paper presented at the annual meeting of the Southern Demographic Association,
Greenville, SC
November 1-2, 2008

ABSTRACT

The effects of population size and growth rate on population forecast accuracy have been well documented. For example, we know that small places generally have larger errors than large places; that errors are generally higher for places with high growth rates than places with low growth rates; and that size of error generally remains more stable over time than does the direction of error. In this paper, we delve more deeply into these relationships using data for 2,482 counties in the United States and expand the analysis to include a third explanatory variable, prior forecast error. We use regression analysis to analyze the effects of these variables on the precision and bias of population forecasts for a number of launch years and forecast horizons between 1900 and 2000. We develop a variety of regression models, some using a single explanatory variable, some using all three, and some using alternative functional forms of the variables. We find that: 1) All three explanatory variables have consistent and statistically significant effects on precision, but the growth rate has the most consistent effect on bias; 2) Alternative nonlinear functional forms of the regression models generally perform better than a linear form for population size and growth rate, but not for prior error; 3) For our measure of precision, the signs and levels of significance for all three explanatory variables remain quite stable over time but the regression coefficients themselves vary substantially; 4) For our measure of bias, the signs and levels of significance remain stable over time only for the growth rate variable; and 5) On balance, the population growth rate contributes slightly more to the discriminatory power of the regression models than does population size and both contribute more than prior forecast error. These findings provide a new perspective on the precision and bias of small-area population forecasts.

INTRODUCTION

Population projections at the state and local levels are used for a wide variety of planning, budgeting, and analytical purposes. Although they are sometimes used as conditional outcomes of a hypothetical set of initial conditions and assumptions, they are used most frequently as forecasts or predictions of future population. The importance of the purposes for which these forecasts are used (e.g., opening a new business, closing an elementary school, enlarging a power plant, or relocating bus stops) makes it essential to evaluate forecast accuracy from as many perspectives as possible and to note any patterns that might be observed.

Many studies have investigated the impact of population size and growth rate on forecast accuracy by analyzing forecast errors within broad size and growth rate categories. Measuring population size in the launch year and growth rate over the base period, these studies have generally found precision (i.e., accuracy regardless of the direction of error) to be positively related to population size and to have a u-shaped relationship with the growth rate, with the smallest errors occurring in places with the smallest growth rates and increasing as growth rates deviate in either direction from these levels (e.g., Keyfitz 1981; Rayer 2008; Smith and Sincich 1992; Stoto 1983; White 1954). They have generally found bias (i.e., accuracy accounting for the direction of error) to be unrelated to population size but positively related to the growth rate (e.g., Isserman 1977; Rayer 2008; Smith 1987; Tayman 1996).

In addition to focusing on broad population size and growth rate categories, most previous studies have been based on a relatively small number of places and time periods and have not delved deeply into the nature of the relationships connecting population size, growth rates, and forecast accuracy. To our knowledge, only one study has approached the issue using regression analysis (Tayman, Schafer, Carter 1998). That study examined 11 aggregated population size categories ranging from 500 to 50,000 in San Diego County.

In this study, we investigate population forecast accuracy from a disaggregate perspective, using a data set encompassing 2,482 counties in the United States from 1900 to 2000. We evaluate several alternative regression models to explain patterns in absolute and algebraic percent forecast errors for individual counties. We extend the analysis to include not only the effects of population size and growth rates on forecast accuracy but a third variable as well, prior forecast error. This variable has been found to be useful in developing empirical prediction intervals (Rayer, Smith, and Tayman 2007; Smith and Sincich 1988) but to our knowledge it has not been evaluated as an independent determinant of population forecast accuracy. The specific questions we address are:

1. What functional forms best describe the relationships between forecast accuracy and population size, growth rate, and prior error?
2. How much of the variation in forecast accuracy can be explained by differences in population size, growth rate, and prior error?
3. What are the relative contributions (strengths) of population size, growth rate, and prior error as determinants of forecast accuracy?
4. How do these relationships vary over time and by length of forecast horizon?

We believe this analysis enhances our understanding of the nature of population growth and the determinants of the precision and bias of small-area population forecasts.

DATA

We conducted our analyses using a data set covering all counties or county equivalents in the United States that did not experience significant boundary changes between 1900 and 2000 (Rayer 2008). This data set included 2,482 counties, 79 percent of the national total. For each county, we collected information on population size in the launch year (the year of the most recent data used to make a forecast), growth rate over the base period (in this study, the 20 years immediately proceeding the launch year), and forecast errors for 10-, 20-, and 30-year horizons. The launch years included in the data set were all decennial census years from 1920 to 1990.

Forecasts for each launch year were derived from five simple extrapolation techniques: linear, exponential, share of growth, shift share, and constant share (Rayer 2008). The forecasts analyzed in this study were calculated as the average of the forecasts from these five techniques after excluding the highest and lowest. They refer solely to total population; no forecasts of age, sex, race, or other demographic characteristics were made. Simple techniques such as these are frequently used for small-area projections and have been found to produce forecasts that are at least as accurate as those produced using more complex or sophisticated techniques (e.g., Long 1995; Murdock et al. 1984; Smith and Sincich 1992; Stoto 1983).

Forecast error was calculated as the percent difference between the population forecasted for a particular year and the population for that year counted in the decennial census. Errors were measured in two ways, one ignoring the direction of error (called “absolute” errors) and the other accounting for the direction of error (called “algebraic” errors). The first is a measure of the precision of population forecasts and the second is a measure of bias.

Table 1 summarizes population size and growth characteristics for counties in the data set. Although mean population size more than doubled between 1920 and 1990, median population size increased by only 27%. The 90th percentile population size grew by 161%, but the 10th percentile population size barely changed. Mean growth rates were higher than median growth rates in every time period and fluctuated substantially over time. In all time periods, between one-quarter and one-half of all counties lost population. For more detailed information on the data set and forecasting techniques, see Rayer (2008).

(Table 1 about here)

ANALYSES

We analyzed the impact of three explanatory variables we believe affect the size and/or direction of population forecast errors: population size in the launch year (Size), population growth rate over the 20-year base period (GR), and forecast error for the time period immediately prior to the launch year (Prior). Although the first two variables have been considered in numerous previous studies, the third has not. Yet, if some places are particularly easy or difficult to forecast accurately due to factors other than population size and growth rate, prior forecast error may serve as a proxy for those factors. For example, the 10-year forecast error for launch year 1950 may be a useful predictor of the 10-year forecast error for launch year 1960.

Table 2 provides a summary of the relationships analyzed in this paper. As seen in the top two panels, the mean absolute percent error (MAPE) has a negative relationship with population size and a u-shaped relationship with the growth rate, whereas the mean algebraic percent error (MALPE) has a positive relationship with both population size and growth rate. A first three relationships are consistent with the findings of most previous studies, but a positive relationship between population size and bias is not. We believe the latter relationship is spurious, caused by a positive correlation between population size and growth rate (Smith, Tayman, and Swanson 2001; Tayman et al 1998). We return to this possibility later in the paper.

(Table 2 about here)

The bottom two panels of Table 2 show the relationship between prior error and forecast accuracy. Prior absolute percent errors display a strong positive relationship with subsequent MAPEs and a weak positive relationship with subsequent MALPEs. Prior algebraic percent errors display a strong u-shaped relationship with subsequent MAPEs and a moderate positive relationship with subsequent MALPEs. We are not aware of any previous studies that have evaluated these relationships.

Table 2 illustrates the approach taken in most studies of the relationships between population characteristics and population forecast accuracy. To examine these relationships more closely, we used regression analysis to investigate the effects of population size, growth rate, and prior error on population forecast errors. We developed several alternative regression models: 1) Each explanatory variable by itself in a simple univariate model; 2) Alternative functional forms of the three single-variable models; 3) The three explanatory variables together in a simple multivariate model; and 4) Alternative functional forms of the multivariate model. We evaluated both absolute and algebraic percent errors as measures of precision and bias, respectively.

Based on theoretical considerations, the results of previous studies, and the data shown in Table 2, we developed the following hypotheses: 1) Increases in population size will improve precision but have no consistent effect on bias; 2) Increases in the absolute value of growth rates will reduce precision; 3) Increases in algebraic growth rates will reduce downward bias in places losing population and raise upward bias in places gaining population (i.e., they will have a positive effect on algebraic percent errors); 4) Increases in the absolute value of prior errors will reduce precision; and 5) Increases in prior algebraic errors will have no consistent effect on bias.

Because we use absolute values of growth rates and prior errors in regressions related to precision and algebraic values in regressions related to bias, we identify the former as GR-Abs and Prior-Abs and the latter as GR-Alg and Prior-Alg. When used as an explanatory variable, prior error is measured using the same number of years as the forecast horizon (e.g., for forecasts covering a 20-year horizon, we used the error for the 20-year forecast ending in the launch year).

Results Using Combined Data

We started by combining all the forecasts into three large data sets. We combined all 10-year forecasts with launch years from 1930 to 1990 into one data set (n = 17,374); all 20-year forecasts with launch years from 1940 to 1980 into a second data set (n = 12,410); and all 30-year forecasts with launch years from 1950 to 1970 into a third data set (n = 7,446). Then, we ran a series of regressions for each of the three data sets.

<u>Simple Univariate Models</u>. Regression coefficients and adjusted R^2 values for the simple univariate models are shown in Table 3. The top panel shows the results for absolute percent errors. With only one exception, all three explanatory variables had the

expected signs and were statistically significant ($\alpha = 0.01$) for all three forecast horizons: increases in population size reduced errors, increases in the absolute value of the growth rate raised errors, and increases in the absolute value of the prior error raised errors. The only exception was the coefficient for population size for 30-year horizons, which was very small and statistically insignificant. As shown by the small adjusted R^2 values, population size and growth rate did not explain much of the variation in the dependent variable; prior error performed somewhat better than the other two variables in this regard.

(Table 3 about here)

There was no clear relationship between the size of the regression coefficient and the length of the forecast horizon for population size, but there was some evidence of a positive relationship for growth rate and a negative relationship for prior error. Increasing the length of the forecast horizon had little effect on the discriminatory power of the model for population size, but had a small positive effect for growth rate and a small negative effect for prior error.

The bottom panel of Table 3 shows the results for algebraic percent errors. Population size had a significant positive effect for all three forecast horizons. This result is inconsistent with our hypothesis that differences in population size have no consistent effect on the direction of forecast errors. We believe this result was spurious, caused by the large sample size, the correlation between population size and other explanatory variables, and the impact of combining forecasts across launch years. We investigate this possibility later in the paper.

The growth rate had a significant positive effect on algebraic errors for all three forecast horizons, as expected. The prior error had a significant positive effect on algebraic errors for the 10-year forecast horizon and significant negative effects for the 20- and 30-year horizons. We believe these mixed results support our hypothesis that the direction of prior errors provides no useful information regarding the direction of future errors. We return to this point when we evaluate the results separately for each launch year. For all three variables, the absolute values of the regression coefficients and the adjusted R^2 values increased with the length of the forecast horizon.

Alternative Single-Variable Models. Our next objective was to explore alternative functional forms to determine whether we could improve the fit of the single-variable models. We used the curve-fit procedure in the SPSS statistical package to identify the models with the fewest parameters and highest adjusted R^2 values. Our selection criteria were that an additional parameter had to be statistically significant and had to add at least 1% to the adjusted R^2. Because the discriminatory power of significance tests tends to decline as sample size increases (Henkel, 1976), these criteria helped us determine whether a parameter made a substantive contribution to the explanation of forecast error. For comparability purposes, we used the same functional form for each launch/horizon combination even if it meant relaxing the adjusted R^2 criterion. The details of the model selection procedures are contained in the Appendix. The regression coefficients and adjusted R^2 values for the optimal alternative models are shown in Table 4.

(Table 4 about here)

The top panel shows the results for absolute percent errors. For population size, the optimal model used the natural log of population size and the square of the natural log. Both variables had significant effects on forecast errors for all three horizons. The natural log had a negative effect and its square had a positive effect, indicating that increases in population size reduced absolute percent errors at a rate that declined as population size increased. Figure 1 shows that the addition of the quadratic term flattens the curve and illustrates the asymptotic nature of the relationship between population size and absolute percent errors. Coefficients for both variables increased as the forecast horizon became longer. Adjusted R^2 values for this model were substantially higher than for the simple univariate model.

(Figure 1 about here)

For growth rates, the optimal model included squared and cubed forms of the variable as well as the variable itself. The first and third terms had positive signs and the second had a negative sign, reflecting a positive but nonlinear relationship between growth rates and forecast errors. As shown in Figure 2, the rate of increase in absolute percent errors tends to decline as growth rates increase, a pattern not evident when analyzing discrete growth rate categories. All three regression coefficients were statistically significant for all three forecast horizons and became larger (in absolute terms) as the horizon became longer. Again, adjusted R^2 values were higher than for the simple univariate model.

(Figure 2 about here)

For prior error, no alternative form could improve on the performance of the simple univariate model; that is, the simplest form proved to be as good as any of the alternative forms we examined, given our selection criteria. Results are therefore identical to those reported previously.

The bottom panel of Table 4 shows the results for algebraic percent errors. The optimal models were the same as they were for absolute percent errors, with one exception: For population size, the optimal model did not contain the square of the natural log, which means the flattening effect for absolute percent errors shown in Figure 1 is not present for algebraic percent errors.

The population size variable had a significant positive effect on algebraic errors for all three horizons. As was true for the simple univariate model, this was counter to our expectations. Regression coefficients and adjusted R^2 values increased monotonically with the length of the forecast horizon. We will return to this finding later in the paper.

All three forms of the growth rate variable were significant for all three horizons, with the first and third terms having positive signs and the second a negative sign. These results indicate that there is a positive but nonlinear relationship between growth rates and algebraic forecast errors. This positive relationship is consistent with our expectations. Regression coefficients for all three growth rate variables increased (in absolute value) with the length of horizon, as did the adjusted R^2 values.

The nature of the relationship between growth rates and algebraic percent errors is illustrated in Figure 3. Forecasts have relatively little bias in areas with low rates of population growth or decline, but become substantially more biased as growth rates deviate in either direction from those low levels. Again, these results demonstrate how nonlinear subtleties in the relationship between growth rate and bias can be obscured when analyzing discrete growth rate categories.

(Figure 3 about here)

Results for prior error were identical to those reported previously, with a significant positive effect for 10-year horizons and significant negative effects for 20- and 30-year horizons.

Adjusted R^2 values were higher for regressions involving absolute percent errors than for those involving algebraic percent errors in seven of the nine variable/horizon year combinations. The only exceptions were for population size and growth rate in the 30-year horizons. This means the explanatory variables generally did a better job explaining variations in the precision of population forecasts than explaining variations in bias.

Simple Multivariate Models. Our analysis thus far has focused on regressions containing a single explanatory variable, sometimes by itself and sometimes with alternative forms of the variable. What happens to the regression coefficients and adjusted R^2 values if we include all three explanatory variables in a single regression? Table 5 shows the results for the simple multivariate model.

(Table 5 about here)

For absolute percent errors, all three variables had the expected signs and prior error was statistically significant for all three forecast horizons, whereas population size and the growth rate were not significant for the 30- and 20-year horizons, respectively. The regression coefficients did not change with the length of the forecast horizon for population size, but generally rose with the horizon for growth rate and declined for prior error. Adjusted R^2 values for the multivariate model were substantially higher than for the simple univariate models for population size and growth rate. For prior error, however, that was not the case, indicating that adding population size and growth rate did not add substantially to the discriminatory power of the simple univariate model. This would seem to imply that much of the impact of prior errors on absolute percent errors in the single-variable model was accounted for by their relationship with population size and growth rate.

For algebraic percent errors, population size and growth rate had positive signs and, except for the growth rate for the 20-year horizon, were statistically significant for all other forecast horizons, whereas prior error had a significant positive effect for the 10-year horizon and significant negative effects for 20- and 30-year horizons. Regression coefficients generally rose with the length of forecast horizon for population size and growth rate, but not for prior error. Adjusted R^2 values increased with the length of horizon and were higher than the values for any of the simple univariate models. Again, the mixed results for prior error are consistent with our hypothesis that the direction of prior errors provides no useful information regarding the direction of future errors. The significant positive effects for population size, however, run counter to our expectations. We return to this point later in the paper.

Alternative Multiple-Variable Models. We constructed multivariate regression models using the alternative functional forms of the explanatory variables described previously. The results are shown in Table 6.

(Table 6 about here)

For absolute percent errors, every variable was statistically significant for every forecast horizon and every one had the same sign as in the single-variable regressions. For every variable except prior error, the absolute value of the regression coefficient increased with the length of the forecast horizon; for prior error, the opposite occurred. For population size and growth rate, this pattern most likely reflects the increase in absolute percent errors associated with longer forecast horizons. The decreasing size of the coefficients for prior error reflects the diminished relevance of prior errors in predicting long-range forecast accuracy. For all three horizons, adjusted R^2 values were substantially higher than they were for the simple multivariate model or any of the univariate models, indicating that the more complex model did a better job explaining variations in precision than the simpler model did.

For algebraic percent errors, every variable was statistically significant for every forecast horizon, except population size for the 10-year horizon, and almost every one had the same sign as in the single-variable regressions. The only exception was the prior algebraic error (Prior-Alg), which had a positive rather than a negative sign for the 30-year horizon. For population size, the regression coefficients increased steadily with the length of the forecast horizon; for prior error, they declined steadily (in absolute terms); and for the growth rate variables, they followed no consistent pattern. Adjusted R^2 values increased with the length of the forecast horizon and were higher than they were in the simple multivariate models.

For both the 10- and 20-year horizons, adjusted R^2 values were substantially higher for regressions involving absolute percent errors than for regressions involving algebraic percent errors. Although this was not true for the 30-year horizon (0.195 vs.

0.153), the two were closer in value than they were in the simple multivariate model. Again, this means the explanatory variables did a better job explaining variations in the precision of population forecasts than in explaining variations in bias.

Relationships among Explanatory Variables. How are the relationships between each explanatory variable and population forecast error impacted by the other variables in the model? To answer this question, we compared coefficients from the alternative single-variable model with coefficients as the other two variables were added sequentially to the regression equation. For example, we examined how the population size coefficients changed after controlling first for the growth rate and then for both the growth rate and prior error.

High intercorrelations among independent variables can confound such comparisons, but the correlations between population size, growth rate, and prior error were relatively low. Ignoring signs, they ranged from 0.015 to 0.157 for 10-year horizons; 0.016 to 0.171 for 20-year horizons; and 0.027 to 0.280 for 30-year horizons. The percent changes in the coefficients for population size, growth rate, and prior error are shown in Tables 7-9, respectively.

(Tables 7-9 about here).

For the regressions relating population size to absolute percent errors, introducing the growth rate reduced the two population size coefficients by 15.8% and 21.5%, respectively, for 10-year horizons; by 14.5% and 19.9% for 20-year horizons; and by 23.8% and 31.4% for 30-year horizons(Table 7). Adding the prior error further accentuated these declines.

As expected, adding the growth rate variable had an even larger impact on the population size coefficients for the algebraic percent errors, with reductions ranging from 36.2% to 69.7%. Adding the prior error had no consistent effect on the results, reducing the coefficient for the 10-year horizon, increasing it for the 20-year horizon, and having little no impact for the 30-year horizon.

For the regressions relating growth rates to absolute percent errors, adding population size increased the coefficients between 8.7% and 22.7% (Table 8). However, adding the prior error reduced the coefficients for all three horizons, actually making the signs negative for 10- and 20-year horizons.

For algebraic percent errors, adding population size to the regressions reduced the magnitude of all but one of the growth rate coefficients. Adding the prior error had an inconsistent effect, sometimes raising the coefficients and sometimes reducing them.

The results presented in Tables 7 and 8 show that adding population size to the growth rate regression generally had less impact on the coefficients than adding the growth rate to the population size regressions. The effects of adding the prior error were mixed, however, sometimes raising the coefficients and sometimes reducing them.

For the regressions relating prior errors to absolute percent errors, adding population size lowered the coefficients by 13.3%, 14.6%, and 28.2% for the 10-, 20-, and 30-year horizons, respectively (Table 9). Adding the growth rate had an even greater impact, raising the reductions to 41.4%, 38.9%, and 66.4% for the three horizons.

There was no consistent impact of adding population size and growth rate to the regressions relating prior errors to algebraic percent errors, however. Adding population size raised the coefficient for the 10-year horizon by 7.3% but lowered it by 11.2% and

53.3% for the 20- and 30-year horizons, respectively. Adding the growth rate dramatically raised the coefficient for the 10-year horizon but dramatically lowered it for the 30-year horizon.

Results by Launch Year

In order to evaluate the stability of results over time, we used the curve-fit procedures and selection criteria described above to identify the optimal model for each combination of launch year and forecast horizon. In most instances, the optimal model was the same as the model found to be optimal in the analyses of the combined data sets. Although this model was sub-optimal in a few instances, we used it across the board in order to provide consistent comparisons for each combination of launch year and forecast horizon. We do not believe enforcing this conformity substantively alters the findings. The Appendix discusses the model selection statistics by launch year.

Absolute Percent Errors. The results for absolute percent errors are shown in Table 10. With respect to signs and levels of significance, the results were quite consistent over launch years and forecast horizons. The log of population size had a negative sign in every instance and was statistically significant 12 of 15 times. The squared term had a positive sign in every instance and was significant 11 of 15 times. The growth rate had a significant positive effect in all launch years for all horizons and the squared and cubed terms had negative and positive coefficients, respectively, in every instance and were statistically significant in most instances. These results are consistent with the results from the combined data sets shown in Table 6. Prior error had a positive effect in every instance except one. These effects were statistically significant in all seven launch years for 10-year horizons and in four of five launch years for 20-year horizons, but were significant in only one launch year for 30-year horizons; this further strengthens the finding from the combined data that the impact of the prior error declines as the forecast horizon becomes longer.

(Table 10 about here)

There were several differences across launch years, however. Regression coefficients varied considerably from one launch year to another for all three forecast horizons. The same was true for the adjusted R^2 values, which ranged from 0.085 to 0.301 for 10-year horizons, from 0.064 to 0.258 for 20-year horizons, and from 0.099 to 0.197 for 30-year horizons. Clearly, the explanatory variables were able to explain a substantially greater proportion of the variance in absolute percent errors in some time periods than in others.

In our discussion of Tables 3-6, we noted several relationships between regression coefficients and the length of the forecast horizon. Were these relationships real or were they spurious, caused by different sets of launch years being included in the combined data sets (1930-1990 for 10-year horizons, 1940-1980 for 20-year horizons, and 1950-1970 for 30-year horizons)? Based on the data shown in Table 10, it appears some of these relationships were real. For each launch year from 1940 to 1980, the absolute values of the regression coefficients increased monotonically with the length of the forecast horizon for population size and growth rate and declined monotonically for prior error; a similar pattern was seen in the combined data. It appears that the impact of population size and growth rate on precision increases with the length of forecast horizon, whereas the impact of prior error declines.

For adjusted R^2 values, however, there was no clear relationship with the length of the forecast horizon; values sometimes rose and sometimes fell as the forecast horizon became longer. It does not appear that the length of the forecast horizon has a consistent effect on the discriminatory power of the model.

Algebraic Percent Errors. Earlier in the paper, we hypothesized that higher growth rates would have a significant positive effect on algebraic percent errors but population size and prior errors would have no consistent effects. These hypotheses were not always supported in analyses based on combined data, but they find more support in the results for individual launch years (Table 11).

(Table 11 about here)

The growth rate had a positive effect on algebraic percent errors in 14 of 15 instances and was statistically significant 10 times. The squared term had a negative coefficient in every instance and was significant nine times; the cubed term had a positive coefficient in all 15 instances and was also significant nine times. These results support the hypothesis that high growth rates during the base period generally lead to an upward bias in population forecasts and high rates of population loss generally lead to a downward bias.

The log of population size sometimes had a significant positive effect on algebraic percent errors, sometimes had a significant negative effect, and sometimes had no significant effect. The same was true for prior error. These results support the hypothesis that there are no consistent relationships between algebraic percent errors and either population size or prior error; that is, neither population size nor prior errors serve as useful indicators of the direction of future forecast errors.

Also, it appears that the positive relationship between adjusted R^2 values and the length of forecast horizon shown in Tables 3-6 was spurious, caused by the specific launch years included in the combined data for each horizon. An analysis of adjusted R^2 values for individual launch years shows that they sometimes increased and sometimes declined with increases in the forecast horizon. Given these results, we do not believe the length of forecast horizon has any consistent impact on the discriminatory power of the model for either absolute or algebraic percent errors.

Impact of Explanatory Variables. Which of the three explanatory variables contributes the most to the discriminatory power of the multivariate models? One way to answer this question is to measure the reduction in adjusted R^2 values that occurs when one of the explanatory variables (including the additional terms for population size and growth rate) is removed from the optimal form of the multivariate model. We interpret this reduction as a measure of that variable's contribution to the model's discriminatory power. The results of this exercise are shown in Tables 12 and 13.

For absolute percent errors, population size variables performed slightly better than growth rate variables when evaluated by launch year, contributing the most in four launch years for 10-year horizons, three launch years for 20-year horizons, and two launch years for 30-year horizons (Table 12). Growth rate variables contributed the most in three, two, and one launch years, respectively. When the results were averaged over all launch years, however, the growth rate variables performed slightly better than the population size variables. For 10-year horizons, removing the growth rate variables reduced the adjusted R^2 by 32% and removing the population size variables reduced it by 27%. For 20-year horizons, removing the growth rate variables reduced the adjusted R^2

by 34% and removing the population size variables reduced it by 24%. For 30-year horizons, the population size and growth rate variables were about equal: Removing each reduced the adjusted R^2 by about 40%. These results suggest that growth rates generally contributed slightly more than population size to the discriminatory power of the model, but the differences were very small.

(Table 12 about here)

The results for prior error are greatly influenced by the functional form of population size and growth rate variables. In the simple univariate models, the adjusted R^2 values for prior error were substantially larger than they were for either population size or growth rate, especially for 10- and 20-year horizons (Table 3). In the optimal forms of the single-variable models, these differences were largely wiped out due to the increased explanatory power of the enhanced models. For 30-year horizons, in fact, the adjusted R^2 value for prior error was smaller than for either population size or growth rate (Table 4). In terms of discriminatory power, it appears that prior error is superior to the other two explanatory variables only when models contain a single form of the explanatory variable.

The results are a bit puzzling for models of algebraic percent errors. As expected, prior error contributed the least to discriminatory power (on average) and removing that variable generally reduced the adjusted R^2 value by a relatively small amount, especially for longer forecast horizons (Table 13). However, although the average reductions were greatest for growth rate variables for both the 10- and 30-year horizons, there were a number of instances in which adjusted R^2 values declined the most when population size variables rather than growth rate variables were removed from the model. We don't have a good explanation for this finding, but it seems to be caused primarily by launch years 1940 and 1950.

(Table 13 about here)

CONCLUSIONS

In this paper, we used regression analysis to analyze the effects of population size, growth rate, and prior error on the precision and bias of population forecasts for a large sample of counties in the United States. We developed a variety of regression models, some using a single explanatory variable, some using all three, and some using alternative functional forms of the variables. We found that: 1) All three explanatory variables had consistent and statistically significant effects on precision, both when modeled by themselves and when combined with each other; 2) The growth rate was the only explanatory variable that had a consistent effect on bias; 3) More complex functional forms of the regression models performed better than simple forms for population size and growth rate, but not for prior error; 4) With respect to precision, the signs and levels of significance of the regression coefficients remained quite stable over time for all three explanatory variables, but the size of the coefficients themselves varied considerably; 5) With respect to bias, the signs and levels of significance remained stable over time only for the growth rate variables; 6) Growth rates contributed slightly more than population size to the discriminatory power of regression models for absolute percent errors, but the differences were very small; and 7) The impact of the prior errors on precision declined substantially as the forecast horizon became longer.

These findings confirm a number of results found previously but also shed light on several issues not considered before. Previous studies provided a general view of the relationships between population size and growth rate and forecast accuracy. The statistical models presented in this study strengthened our understanding of the nature, strength, and temporal stability of these relationships and incorporated a variable not generally considered; namely, prior forecast error. These models quantified the effects of changes in population size, growth rate, and prior error on the precision and bias of county population forecasts. They also uncovered several subtleties in the relationships among these variables that were masked in previous analyses.

The diminishing impact of increases in population size on improvements in forecast precision is quite evident in the asymptotic relationship. The relationship between population size and bias is also asymptotic and was statistically significant in several instances. Changes in the direction and strength of this relationship over time, however, are indicative of the unpredictable nature of this relationship. In some instances, bias was negative for smaller areas and became increasingly more positive with increasing population size and in other instances the opposite pattern occurred.

Increases in absolute growth rates had a diminishing impact on changes in forecast precision. The cubic relationship between forecast bias and growth rate was more complex than the presumed linear form. Bias tended to be relatively low in stable or slowly changing areas, but downward bias increased with larger rates of population decrease and upward bias increased with larger rates of population growth.

The impact of prior absolute percent error on forecast precision was consistently negative (i.e., increases in prior errors reduced precision) but the impact of prior algebraic percent error was inconsistent. Patterns in the adjusted R^2 and coefficient values indicated that prior error would be more effective in predicting errors for shorter forecast horizons than longer horizons.

Growth rate and population size had more explanatory power than prior error in multivariate regressions and, on balance, the growth rate had slightly more explanatory power than population size. Furthermore, the growth rate was clearly more influential than the other two explanatory variables in terms of its impact on their coefficient values. In particular, the effect of population size on bias was reduced substantially when the growth rate was added into the model.

These findings provide a new perspective on the precision and bias of subnational population forecasts. However, the relatively low adjusted R^2 values found in many of the models imply that population size, growth rate, and prior error by themselves can explain only a small amount of the county-to-county variation in forecast errors. While we examined the individual effects of population size and growth rate, their joint effects or interactions may also have an impact on forecast precision and bias. For example, it seems reasonable that the relationship between population size and forecast errors might vary with differences in the growth rate, with stable areas showing weaker relationships than rapidly changing areas. Another factor not considered here is the impact of differences in geographic region, which may serve as a proxy for subnational factors that influence population forecast errors independently of differences in population size, growth rate, and prior error. Clearly, many other factors are at work and much remains to be done to improve our understanding of the determinants of the accuracy of small-area population forecasts.

REFERENCES

Henkel, R. 1976. *Tests of Significance*. Quantitative Applications in the Social Sciences #4. Sage Publications, Beverly Hills, CA.

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

Keyfitz, N. 1981. The limits of population forecasting, *Population and Development Review*, 7, 579-593.

Long, J. F. 1995. Complexity, accuracy, and utility of official population projections, *Mathematical Population Studies*, 5, 203-216.

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

Rayer, S. 2008. Population forecast errors: A primer for planners, *Journal of Planning Education and Research*, 27, 417-430.

Rayer, S., S. K. Smith, and J. Tayman, J. 2007. Empirical prediction intervals for county population forecasts. Poster presented at the annual meeting of the Population Association of America, New York, NY.

Smith, S. K. 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 forecasts errors, *Demography*, 25, 461-474.

Smith, S. K. and T. Sincich. 1992. Evaluating the forecast accuracy and bias of alternative population projections for states, *International Journal of Forecasting*, 8, 495-508.

Smith, S. K., Tayman, J. and Swanson, D. 2001. *State and local population projections: Methodology and analysis.* Kluwer Academic/Plenum Publishers, New York.

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

Tayman, J. 1996. The accuracy of small area population forecasts based on a spatial interaction land use modeling system, *Journal of the American Planning Association*, 62, 85-98.

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 interval for subcounty areas, *Population Research and Policy Review*, 17(1): 1-20.

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

Appendix: Selecting the Functional Form

Combined Data Set

A wide range of functional forms were evaluated to determine the best fitting function to explain the absolute percent error (APE) and algebraic percent error (ALPE). Six univariate relationships were analyzed: population size-APE, growth rate-APE, prior error-APE, population size-ALPE, growth rate-ALPE, and prior error-ALPE. The goal was to seek the most parsimonious function whose parameters were both statistically significant and added at least 1% to the unadjusted R^2. For comparability, our aim was to use the same function, which could vary across the six relationships, for each launch/horizon combination even if it meant relaxing the R^2 criterion. These functions were selected using the combined data sets analyzed separately for the 10-, 20-, and 30-year forecast horizons. The following functions were investigated:

Equation	Name
1. $y = b0 + (b1 * t)$	Linear
2. $y = b0 + (b1 * (\ln(t))$	Logarithmic
3. $y = b0 + (b1 * 1/t)$	Inverse
4. $y = b0 + (b1 * t) + (b2 * t^2)$	Quadratic[a]
5. $y = b0 + (b1 * t) + (b2 * t^2) + (b3 * t^3)$	Cubic[a]
6. $\ln(y) = \ln(b0) + (b1 * \ln(t))$	Power
7. $\ln(y) = \ln(b0) + (\ln(b1) * t)$	Compound
8. $\ln(y) = b0 + (b1 / t)$	S
9. $\ln(y) = b0 + (b1 * t)$	Growth
10. $\ln(y) = \ln(b0) + (b1 * t)$	Exponential

[a] The quadratic and cubic functions were evaluated for Size, ln(Size), inverse (1/Size), GR-Abs, and GR-Alg.

The functions estimated depended on the signs of the independent and dependent variables. Functions based on the natural log of the dependent variable could not be used to predict the ALPE because the natural log of a negative number is undefined. For the same reason, the logarithmic function was also not viable in equations using the GR-Alg and Prior-Alg as independent variables to predict the ALPE. For the APE equations, all functions were tested with two exceptions; the logarithmic and power functions could not be estimated using GR-Alg as the independent variable. The specific relationships evaluated for the APE and ALPE used the following independent variables: Size, ln(Size), GR-Abs, GR-Alg, and either Prior-Abs or Prior-Alg.

<u>Population Size and APE.</u> The selected function was the quadratic based on the ln(Size). Both the selected function and (1/Size) improved upon the linear specification and explained more variation than any function based on the ln(APE). The explained variances were similar between ln(Size) and (1/Size) and varied by less than one percent for 7 of the 9 comparisons (3 horizon years and linear, quadratic, and cubic functions). The R^2 of ln(Size), however, was higher than (1/Size) for the quadratic and cubic functions for 30-yr horizons by 1.7% and 1.5%, respectively. In moving from the linear to quadratic form, the R^2 increase for ln(Size) ranged from 1.0% to 3.4% across horizon years, but hardly changed when the cubic term was added. The quadratic function based on ln(Size) explained between 1% and 7% more variance across horizon years than the quadratic function base on Size.

<u>Growth Rate and APE.</u> The selected function was the cubic based on the GR-Abs. For 10- and 20-year horizons, the R^2 increased by 3.8% and 2.0%, respectively after adding the quadratic term, but for the 30-year horizon the increase was below the

threshold at 0.8%. Adding the cubic term further increased the R^2 by 1.6% and 1.3% for the 10- and 30-year horizons, respectively, but for the 20-year horizon the increase was just below the threshold at 0.9%. Using GR-Abs improved upon the equations based on GR-Alg, with R^2 increases between 2.3% and 3.0% higher across horizon years for the cubic function. Additionally, the signs of the coefficients using GR-Abs were more stable over time than the signs using GR-Alg.

Prior Error and APE. The selected function was linear. The quadratic and cubic functions had an R^2 close to the linear function and far higher than the other functions. Adding the quadratic and cubic parameters had little effect on the R^2, with changes in R^2 ranging from 0.0% to only 0.3%.

Population Size and ALPE. The selected function was linear based on the ln(Size). For 10- and 20-year horizons, the differences in R^2 between Size and ln(Size) were less than the 1% threshold, but for the 30-year horizon the R^2 for ln(Size) was substantially higher (8.8%). The addition of the quadratic and cubic parameters to ln(Size) had virtually no impact on the R^2.

Growth Rate and ALPE. The selected function was cubic based on GR-Alg. The quadratic parameter increased the R^2 between 1.1% and 2.8% across forecast horizons, compared with the linear function. The impact on R^2 of the cubic parameter was not as consistent. For 10- and 20-year horizons, the cubic term increased the R^2 by only 0.4% and 0.6%, respectively, but for the 30-year horizon the cubic term raised the R^2 by 3.2%. In predicting the ALPE, GR-Alg outperformed GR-Abs, opposite of the relationship between growth rate and APE. The R^2 of the cubic function based on GR-Alg was between 1.6% and 14.8% higher than GR-Abs across horizon years.

Prior Error and ALPE. The selected function was linear, the same function found for the relationship between prior error and APE. Adding the quadratic term did not change the R^2 for the 10- and 30-year horizons and only increased it by 0.5% in the 20-year horizon. A similar pattern was seen with the addition of the cubic parameter.

Individual Launch Years

Using the same criteria and functional forms discussed above, we evaluated, the six relationships for each of the 15 launch/horizon year combinations. These combinations are: 1930 to1990 launch years for the 10-year horizon, 1940 to1980 launch years for the 20-year horizon, and 1950 to 1970 launch years for the 30-year horizon.

Population Size and APE. The quadratic function based on ln(Size) was the optimal function for 12 of the 15 launch/horizon year combinations. For 3 combinations, the quadratic term added little to the explained variance (between 0.2% and 0.4%).

Growth Rate and APE. The cubic function based on GR-Abs was the optimal function for 10 of the 15 launch/horizon year combinations. For 3 combinations a quadratic model would have met the criteria and a linear model would have been sufficient for the other 2 combinations.

Prior Error and APE. The linear function was the optimal model in 11 of the 15 launch/horizon year combinations. In the 4 other combinations, a quadratic model would have met the criteria, adding 1.2% to 2.3% to the explained variance of the linear model.

Population Size and ALPE. The linear function based on ln(Size) was the optimal model in 11 of the 15 launch/horizon year combinations. In the 4 other combinations, a quadratic model would have met the criteria, adding 1.0% to 1.7% to the explained variance of the linear function.

Growth Rate and ALPE. The cubic function based on GR-Alg was the optimal model for 9 of the 15 launch year combinations. In one instance a quadratic model would have been sufficient, and a linear model for the other 5 launch/horizon year combinations. In 3 of these 5 combinations, however, the R^2 did not exceed 0.3% for any functional form.

Prior Error and ALPE. The linear function was the optimal model in less than half of the launch/horizon year combinations (7 of 15), the least consistent of any relationship. In 2 instances a cubic model would have met the criteria adding 1.3% and 4.2% to the explained variance of the linear model. This rather large increase in the explained variance was caused by eight outlying cases. In the other 6 launch/horizon year combinations, a quadratic function would have met the criteria, adding between 1% and 3% to the explained variance of a linear model.

Table 1. County Population Size and Growth Characteristics, 1920-1990

Size	1920	1930	1940	1950	1960	1970	1980	1990
Mean	35,044	40,615	43,336	49,453	58,502	66,122	72,950	79,054
Median	18,462	18,570	19,285	19,269	19,236	19,454	22,651	23,376
10 Percentile	5,876	6,483	6,409	6,124	5,780	5,572	6,052	5,820
90 Percentile	57,065	65,870	72,618	84,323	99,673	115,384	136,897	149,116
Growth Rate*								
Mean	70.9	26.7	30.9	11.4	13.2	13.6	24.0	22.0
Median	13.1	5.7	6.5	4.8	0.3	2.6	14.6	12.9
10 Percentile	-13.4	-16.9	-13.2	-20.7	-28.5	-24.6	-14.3	-13.5
90 Percentile	105.1	62.3	51.5	45.2	66.4	58.5	68.9	65.0
% Negative	32.3	39.6	36.5	40.7	49.7	45.8	25.6	28.6

* Percentage change over previous 20 years.

Table 2. Selected Error Measures for Counties by Population Size, Growth Rate, and Prior Error[a]

Size	MAPE	MALPE	% Positive	Sample Size
< 7,500	14.2	-2.5	43.0	2,495
7,500 to 14,999	10.8	-1.5	45.1	3,981
15,000 to 29,999	9.0	-1.1	44.8	4,931
30,000 to 99,999	8.3	-0.5	47.7	4,233
100,000+	8.3	0.3	51.3	1,734

Growth Rate*	MAPE	MALPE	% Positive	Sample Size
< -10%	11.0	-6.3	31.2	3,769
-10.0 to 9.9%	8.0	-0.3	47.0	5,705
10.0 to 24.9%	8.6	0.8	53.5	3,217
25.0 to 74.9%	10.5	1.0	52.8	3,467
75.0%+	17.0	0.4	47.0	1,216

Prior Abs. % Error	MAPE	MALPE	% Positive	Sample Size
< 2.0%	8.0	-1.9	43.5	2,486
2.0 to 3.9%	8.1	-1.0	44.9	2,376
4.0 to 7.9%	8.5	-1.5	43.4	4,072
8.0 to 14.9%	9.7	-0.8	46.8	4,387
15.0 to 24.9%	11.9	-0.8	50.2	2,561
25.0+%	17.0	-0.5	49.1	1,492

Prior Alg. % Error	MAPE	MALPE	% Positive	Sample Size
< -15.0%	15.0	-2.3	47.6	2,187
-15.0 to -8.0%	10.5	-1.6	47.9	2,188
-7.9 to 0.0%	8.7	-1.7	43.6	4,388
0.0% to 7.9%	7.9	-1.2	44.0	4,546
8.0% to 14.9%	8.9	-0.1	45.6	2,199
15.0+%	12.4	1.3	52.3	1,866

[a] All 10-year horizon forecasts from launch years 1930 to 1990

* Percentage change 20 years prior to launch year

Table 3. Univariate Regression Models

Absolute Percent Errors

Ind. Var.	Horizon 10	Horizon 20	Horizon 30
Size[a]			
Coeff[b]	-0.002*	-0.002*	< 0.001
Adj R^2	0.002	0.001	0.000
GR-Abs			
Coeff[b]	0.007*	0.006*	0.101*
Adj R^2	0.009	0.003	0.038
Prior-Abs			
Coeff[b]	0.256*	0.198*	0.131*
Adj R^2	0.079	0.054	0.045

Algebraic Percent Errors

Ind. Var.	Horizon 10	Horizon 20	Horizon 30
Size[a]			
Coeff[b]	0.002*	0.008*	0.030*
Adj R^2	0.001	0.005	0.037
GR-Alg			
Coeff[b]	0.005*	0.009*	0.218*
Adj R^2	0.002	0.003	0.108
Prior-Alg			
Coeff[b]	0.041*	-0.125*	-0.137*
Adj R^2	0.002	0.020	0.040

[a] Size measured in thousands of persons
[b] Unstandardized regression coefficient
* Significant at 0.01

Table 4. Alternative Functional Forms of Single-Variable Regression Models

Absolute Percent Errors

Ind. Var.[a]	Horizon 10	20	30
Size			
Ln Size	-12.479*	-19.299*	-35.686*
$(\text{Ln Size})^2$	0.530*	0.826*	1.569*
Adj R^2	0.050	0.038	0.067
GR-Abs			
GR-Abs	0.056*	0.072*	0.281*
$(\text{GR-Abs})^2$	-0.000017*	-0.000022*	-0.000964*
$(\text{GR-Abs})^3$	0.000000001*	0.000000002*	0.000000785*
Adj R^2	0.062	0.031	0.060
Prior-Abs			
Prior-Abs	0.256*	0.198*	0.131*
Adj R^2	0.079	0.054	0.045

Algebraic Percent Errors

Ind. Var.[a]	Horizon 10	20	30
Size			
Ln Size	0.673*	2.431*	9.364*
Adj R^2	0.003	0.013	0.125
GR-Alg			
GR-Alg	0.038*	0.076*	0.471*
$(\text{GR-Alg})^2$	-0.000012*	-0.000024*	-0.001772*
$(\text{GR-Alg})^3$	0.0000000001*	0.0000000018*	0.000001486*
Adj R^2	0.018	0.020	0.169
Prior-Alg			
Prior-Alg	0.041*	-0.125*	-0.137*
Adj R^2	0.002	0.020	0.040

[a] Unstandardized regression coefficients

* Significant at 0.01

Table 5. Simple Multivariate Regression Model

Absolute Percent Errors

	Horizon		
Ind. Var.[a]	10	20	30
Size[b]	-0.002*	-0.002*	-0.002
GR-Abs	0.003*	0.001	0.085*
Prior-Abs	0.248*	0.196*	0.111*
Adj. R^2	0.082	0.054	0.070

Algebraic Percent Errors

	Horizon		
Ind. Var.[a]	10	20	30
Size[b]	0.002*	0.008*	0.020*
GR-Alg	0.006*	0.003	0.187*
Prior-Alg	0.048*	-0.119*	-0.027*
Adj. R^2	0.006	0.024	0.126

[a] Unstandardized regression coefficients
[b] Size measured in thousands of persons
* Significant at 0.01

Table 6. Alternative Functional Forms of Multivariate Regression Models

Absolute Percent Errors

Ind. Var.[a]	Horizon 10	Horizon 20	Horizon 30
Ln Size	-8.749*	-13.688*	-25.414*
$(\text{Ln Size})^2$	0.345*	0.550*	1.009*
GR-Abs	0.049*	0.063*	0.320*
$(\text{GR-Abs})^2$	-0.000016*	-0.000021*	-0.001040*
$(\text{GR-Abs})^3$	0.000000001*	0.000000002*	0.000000814*
Prior-Abs	0.150*	0.121*	0.044*
Adj. R^2	0.147	0.094	0.153

Algebraic Percent Errors

Ind. Var.[a]	Horizon 10	Horizon 20	Horizon 30
Ln Size	0.138	1.628*	4.599*
GR-Alg	0.047*	0.034*	0.419*
$(\text{GR-Alg})^2$	-0.000015*	-0.000011*	-0.00158*
$(\text{GR-Alg})^3$	0.000000001*	0.0000000001*	0.00000135*
Prior-Alg	0.094*	-0.082*	0.060*
Adj. R^2	0.027	0.031	0.195

[a] Unstandardized regression coefficients
* Significant at 0.01

Table 7. Impact of Growth Rate and Prior Error on Population Size Coefficients[a]

Absolute Percent Errors

	Add Growth Rate		Add Growth Rate and Prior Error	
Forecast Horizon	ln(Size)	$\ln(\text{Size})^2$	ln(Size)	$\ln(\text{Size})^2$
10	-15.8%	-21.5%	-29.9%	-34.9%
20	-14.5%	-19.9%	-29.1%	-33.4%
30	-23.8%	-31.4%	-28.8%	-35.7%

Algebraic Percent Errors

Forecast Horizon	ln(Size)	ln(Size)
10	-69.7%	-79.5%
20	-36.2%	-33.0%
30	-49.7%	-50.9%

[a] Percent change from alternative single-variable model coefficients

Table 8. Impact of Size and Prior Error on Growth Rate Coefficients[a]

Absolute Percent Errors

	Add Size			Add Size and Prior Error		
Forecast Horizon	GR-Abs	$(GR\text{-}Abs)^2$	$(GR\text{-}Abs)^3$	GR-Abs	$(GR\text{-}Abs)^2$	$(GR\text{-}Abs)^3$
10	10.7%	17.6%	16.7%	-12.5%	-5.9%	-1.6%
20	16.7%	22.7%	21.1%	-12.5%	-4.5%	-5.0%
30	21.4%	13.0%	8.7%	13.9%	7.4%	3.7%

Algebraic Percent Errors

Forecast Horizon	GR-Alg	$(GR\text{-}Alg)^2$	$(GR\text{-}Alg)^3$	GR-Alg	$(GR\text{-}Alg)^2$	$(GR\text{-}Alg)^3$
10	-2.6%	0.0%	-5.1%	23.7%	25.0%	18.4%
20	-18.4%	-20.8%	-20.6%	-55.3%	-54.2%	-53.7%
30	-24.4%	-22.2%	-20.1%	-11.0%	-10.6%	-9.4%

* Percent change from alternative single variable model coefficients

Table 9. Impact of Size and Growth Rate on Prior Error Coefficients[a]

Absolute Percent Errors

	Add Size	Add Size and Growth Rate
Forecast Horizon	Prior-Abs.	Prior-Abs.
10	-13.3%	-41.4%
20	-14.6%	-38.9%
30	-28.2%	-66.4%

Algebraic Percent Errors

Forecast Horizon	Prior-Alg.	Prior-Alg.
10	7.3%	129.3%
20	-11.2%	-34.4%
30	-53.3%	-143.8% [b]

* Percent change from alternative single variable model coefficients

[b] Coefficient changed sign

Table 10. Regression Results for Multivariate Model, by Launch Year and Horizon, Absolute Percent Errors

10-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	-25.583*	-12.740*	-3.995	-6.065*	-11.319*	-1.594	-5.814*
(Ln Size)^2	1.076*	0.524*	0.182	0.185*	0.442*	0.049	0.228*
GR-Abs	0.091*	0.037*	0.122*	0.110*	0.077*	0.055*	0.103*
(GR-Abs)^2	-0.000046*	-0.000012*	-0.000267*	-0.000326*	-0.000280	-0.000155	-0.000376*
(GR-Abs)^3	0.000000005*	0.0000000001*	0.000000173	0.000000308*	0.000000294	0.000000137	0.000000484*
Prior-Abs	0.084*	0.122*	0.117*	0.189*	0.258*	0.096*	0.098*
Adj R^2	0.301	0.129	0.107	0.208	0.183	0.085	0.197

20-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	---	-17.036*	-6.208	-23.240*	-19.751*	-4.841*	---
(Ln Size)^2	---	0.680*	0.238	0.922*	0.835*	0.186	---
GR-Abs	---	0.057*	0.229*	0.125*	0.159*	0.124*	---
(GR-Abs)^2	---	-0.000019*	-0.000541*	-0.000468*	-0.000847*	-0.000378*	---
(GR-Abs)^3	---	0.000000001*	0.000000341*	0.000000430*	0.000001093*	0.000000372	---
Prior-Abs	---	0.067*	-0.005	0.223*	0.266*	0.077*	---
Adj R^2	---	0.064	0.067	0.258	0.196	0.114	---

30-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	---	---	-17.751*	-34.278*	-19.898*	---	
(Ln Size)^2	---	---	0.687*	1.359*	0.798*	---	---
GR-Abs	---	---	0.292*	0.375*	0.244*	---	---
(GR-Abs)^2	---	---	-0.000781*	-0.001297*	-0.001119*	---	---
(GR-Abs)^3	---	---	0.000000525*	0.000001085*	0.000001490	---	---
Prior-Abs	---	---	0.009	0.022	0.149*	---	---
Adj R^2	---	---	0.099	0.197	0.181	---	---

[a] Unstandardized regression coefficients

Table 11. Regression Results for Multivariate Model, by Launch Year and Horizon, Algebraic Percent Errors

10-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	0.696	-3.991*	-1.100*	-0.061	3.079*	-0.646*	0.631*
GR-Alg	0.173*	0.004	0.014	0.185*	0.122*	0.062*	-0.006
$(GR\text{-}Alg)^2$	-0.000086*	-0.000002	-0.000151	-0.000711*	-0.000687*	-0.000700*	-0.000168
$(GR\text{-}Alg)^3$	0.000000009*	.0000000002	0.000000147	0.000000669*	0.000001107*	0.000000981*	0.000000210
Prior- Alg	0.121*	-0.056*	0.307*	0.152*	0.342*	0.075*	0.083*
Adj R^2	0.261	0.074	0.135	0.134	0.337	0.057	0.054

20-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	---	-7.876*	-2.258*	3.325*	6.375*	1.399*	---
GR- Alg	---	-0.00023	0.048	0.265*	0.116*	0.067*	---
$(GR\text{-}Alg)^2$	---	-0.000002	-0.000208	-0.001157*	-0.001166*	-0.000979*	---
$(GR\text{-}Alg)^3$	---	0.0000000002	0.000000196	0.000001061*	0.000002025*	0.000001404*	---
Prior- Alg	---	-0.084*	0.060*	-0.089*	-0.176*	0.012	---
Adj R^2	---	0.068	0.010	0.273	0.267	0.057	---

30-Year Horizon

Ind. Var.[a]	1930	1940	1950	1960	1970	1980	1990
Ln Size	---	---	1.715*	3.240*	7.648*	---	---
GR- Alg	---	---	0.099*	0.695*	0.276*	---	---
$(GR\text{-}Alg)^2$	---	---	-0.000398	-0.002672*	-0.001799*	---	---
$(GR\text{-}Alg)^3$	---	---	0.000000357	0.000002270*	0.000002873*	---	---
Prior- Alg	---	---	0.012	0.160*	-0.091*	---	---
Adj R^2	---	---	0.013	0.341	0.317	---	---

[a] Unstandardized regression coefficients

Table 12. Percent Reduction in Adjusted R^2 Values after an Explanatory Variable is Removed from the Multivariate Model, Absolute Percent Errors

		Length of Horizon		
Launch Year	Variable Removed	10	20	30
1930	Size	37.5	---	---
	GR	32.6	---	---
	Prior	2.0	---	---
1940	Size	28.7	37.5	---
	GR	24.0	32.8	---
	Prior	10.9	6.3	---
1950	Size	1.9	10.4	40.4
	GR	42.1	83.6	61.6
	Prior	12.1	-1.5[a]	0.0
1960	Size	33.7	31.8	54.3
	GR	28.4	5.8	42.6
	Prior	12.5	19.8	0.5
1970	Size	40.4	30.6	24.9
	GR	7.7	7.7	14.4
	Prior	28.4	30.1	22.1
1980	Size	11.8	10.5	---
	GR	30.6	41.2	---
	Prior	18.8	9.6	---
1990	Size	31.0	---	---
	GR	60.9	---	---
	Prior	5.1	---	---
Average	Size	26.4	24.2	39.9
	GR	32.3	34.2	39.5
	Prior	12.8	12.9	7.5

[a] Adjusted R^2 greater than the full 3-variable multivariate model

Table 13. Percent Reduction in Adjusted R^2 Values after an Explanatory Variable is Removed from the Multivariate Model, Algebraic Percent Errors

Launch Year	Variable Removed	Length of Horizon 10	20	30
1930	Size	0.4	---	---
	GR	88.1	---	---
	Prior	5.0	---	---
1940	Size	93.2	98.5	---
	GR	1.4	-1.5[a]	---
	Prior	5.4	7.4	---
1950	Size	3.7	80.0	33.3
	GR	0.0	0.0	33.3
	Prior	51.9	20.0	13.3
1960	Size	0.0	5.9	2.3
	GR	70.1	13.6	50.1
	Prior	10.4	1.5	9.1
1970	Size	14.5	27.3	18.9
	GR	10.4	4.5	5.4
	Prior	30.3	4.9	1.9
1980	Size	7.0	12.3	---
	GR	71.9	64.9	---
	Prior	12.3	0.0	---
1990	Size	18.5	---	---
	GR	38.9	---	---
	Prior	14.8	---	---
Average	Size	19.6	44.8	18.2
	GR	40.1	16.3	29.6
	Prior	18.6	6.7	8.1

[a] Adjusted R^2 greater than the full 3-variable multivariate model

Figure 1. Natural Log Functions- Population Size

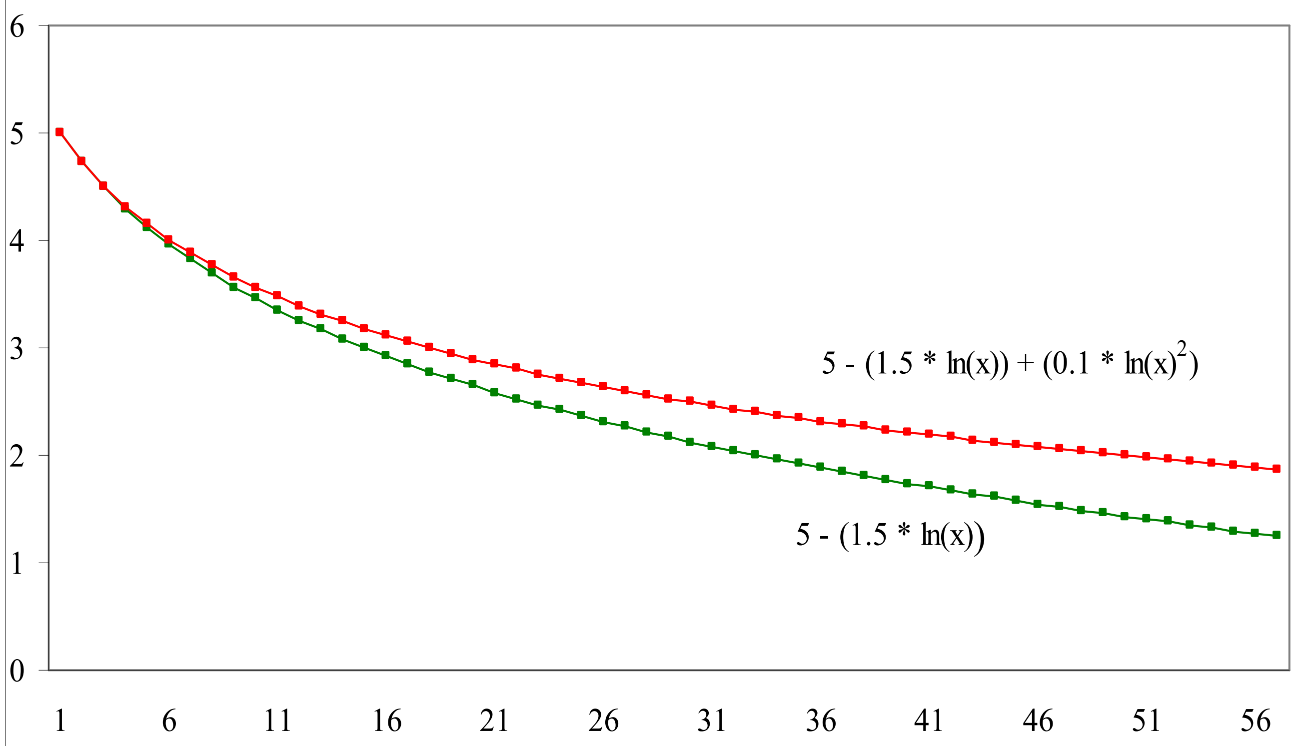

Figure 2. Cubic Function- Absolute Value of Growth Rate

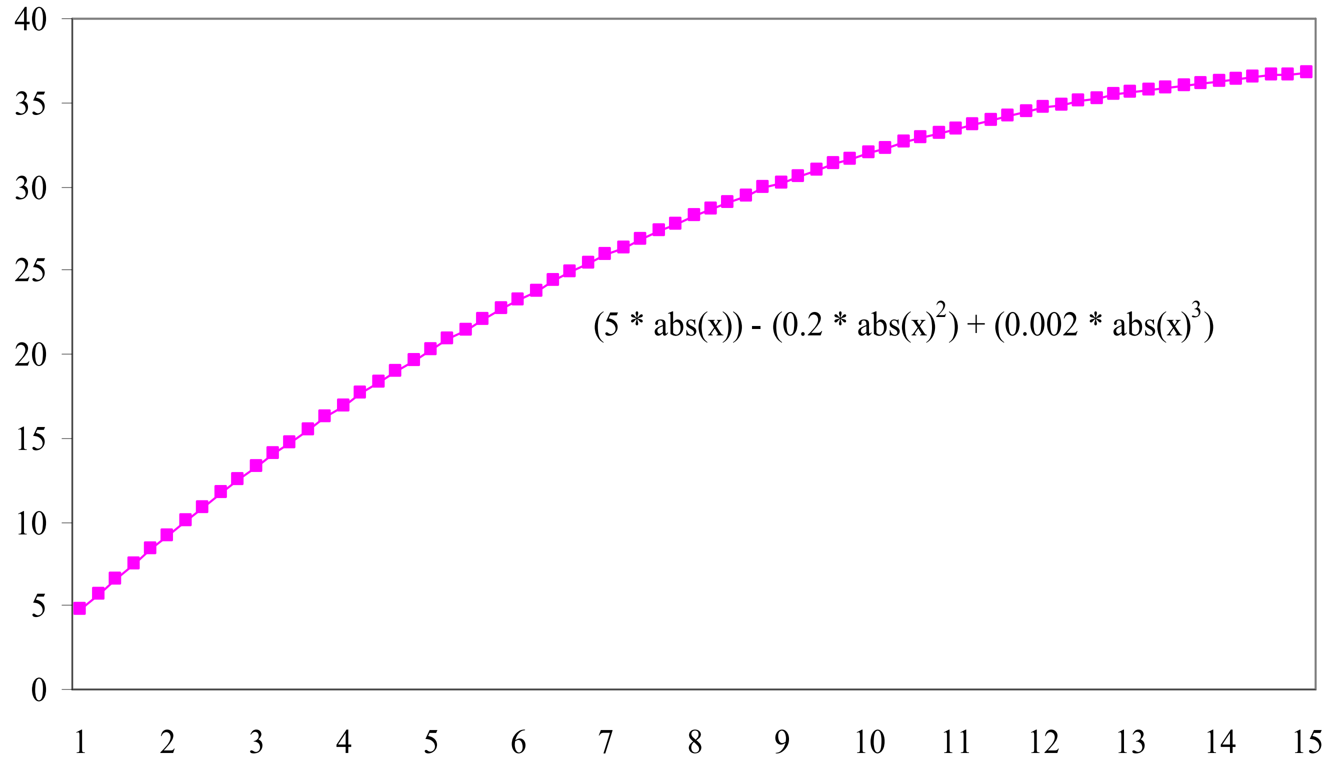

Figure 3. Cubic Function- Algebraic Growth Rate