MAPE-R: A RESCALED MEASURE OF ACCURACY FOR CROSS-SECTIONAL FORECASTS

David A. Swanson
Department of Sociology
University of California Riverside
Riverside, CA 92521 USA
(email: David.swanson@ucr.edu)

Jeff Tayman
Department of Economics
University of California San Diego
La Jolla, CA 92093

T. M. Bryan
McKibben Demographic Research
PO Box 2921
Rock Hill, South Carolina 29732

MAPE-R: A RESCALED MEASURE OF ACCURACY
FOR CROSS-SECTIONAL FORECASTS

Abstract

Accurately measuring a population and its attributes at past, present, and future points in time has been of great interest to demographers. Within discussions of forecast accuracy, demographers have often been criticized for their inaccurate prognostications of the future. Discussions of methods and data are usually at the center of these criticisms, along with suggestions for providing an idea of forecast uncertainty. The measures used to evaluate the accuracy of forecasts also have received attention and while accuracy is not the only criteria advocated for evaluating demographic forecasts, it is generally acknowledged to be the most important. In this paper, we continue the discussion of measures of forecast accuracy by concentrating on a rescaled version of a measure that is arguably the one used most often, Mean Absolute Percent Error (MAPE). The rescaled version, MAPE-R, has not had the benefit of a major empirical test, which is the central focus of this paper. We do this by comparing 10-year population forecasts for U.S. counties to 2000 census counts. We find that the MAPE-R offers a significantly more meaningful representation of average error than MAPE in the presence of outlying errors and provide guidelines for its implementation.

Keywords: MAPE, MAPE-R, national county test, forecast accuracy

1. Introduction

Accurately measuring a population and its attributes at a present point in time has been the subject of a great deal of demographic and related literature (Anderson and Fienberg 2001; Brass et al. 1968; Choldin 1994; Desrosières 1998; Dillman Smyth and Christian 2008; Hough and Swanson 2006; Howe 1999; Hull 2007; Skerry 2000), as has been the case for a past point in time (Caldwell and Caldwell 2003; Chamberlain 2006; Preston et al. 2003; Schmitt 1977; Thornton 1987; and Zerubavel 2003). Therefore, it is not surprising that attempting to accurately measure the future also has been a subject of interest for demographers (Ahlburg 1992; Armstrong and Collopy 1992; Bell and Skinner 1992; Booth 2004; Bulatao 2001; Campbell 2002; de Gans 1999; Isserman 1977; Isserman and Fisher 1984; Keilman 1999; Keyfitz 1972 1981 1982 1987; Lee 2000; Moen 1984; Mulder 2002; Pflaumer 1988; Pittenger 1978; Romaniuc 1990 1992; Sanderson 1998; Smith 1987; Smith and Sincich 1988 1990 1991 1992; Swanson and Tayman 1995; Wilson 2007).

Within these discussions, demographers have often been criticized for their inaccurate forecasts (Ahlburg, Lutz, and Vaupel 1998; Bauer 1981: 50; Dorn 1950; Hajnal 1955; National Research Council 1980, 2000; Simon 1996). Discussions of methods and data are usually at the center of these criticisms, along with suggestions for providing an idea of forecast uncertainty (Alho, Jensen, and Lassila 2008; Alho and Spencer 1985 2005; Cohen 1986; Swanson 2008; Swanson and Beck 1994; Stoto 1983; Sykes 1969). However, the measures used to evaluate the accuracy of forecasts also have received attention (Ahlburg 1992; Armstrong and Fildes 1995; Clements and Hendry 1993; Coleman and Swanson 2007; Long 1995; Mulder 2002; Swanson and Tayman 1995; Tayman, Schafer, and Carter 1998; Tayman and Swanson 1999), and while accuracy is not the only criteria advocated for evaluating demographic forecasts, it is generally acknowledged to be the most important (Ahlburg Lutz and Vaupel 1998; Smith, Tayman, and Swanson 2001: 280-299; Yokum and Armstrong 1995).

In this paper, we continue the discussion of measures of forecast accuracy by concentrating on a rescaled version of the measure that is arguably the one used most often, Mean Absolute Percent Error (MAPE). The rescaled version, MAPE-R, was introduced by Tayman, Swanson, and Barr (1999), given a limited empirical test by Swanson, Tayman, and Barr (2000), and conceptually and computationally refined by Coleman and Swanson (2007). MAPE-R is based on a power transformation of the error distribution underlying the MAPE. It is designed to address the impact of outlying errors on the MAPE, which can overstate the error represented by "most" of the observations, while still preserving the valuable statistical properties of an average. However, MAPE-R has not had the benefit of a major empirical test, which is the central focus of this paper, along with providing criteria for MAPE-R's implementation. The initial investigations of MAPE-R included empirical illustrations in the form of case studies for estimates (Swanson, Tayman, and Barr, 2000) and for forecasts (Tayman Swanson, and Barr 1999). However, the case study data were not intended to provide a definitive empirical portrait of MAPE-R, its features and characteristics

2. Measures of Forecast Accuracy

Swanson and Stephan (2004: 770) define a population forecast as "… an approximation of the future size of the population for a given area, often including its composition and distribution. A population forecast usually is one of a set of projections selected as the most likely representation of the future." Using this definition of a forecast, population forecast error is then the difference between the observed and the forecasted population at a designated point in forecast period; that is E = (F – O). This follows a long-standing tradition of using the "ex-post facto" perspective in examining forecast error, where the error of a forecast is evaluated relative to what was subsequently observed, typically a census-based benchmark (Campbell 2002; Mulder 2002). Forecast errors can be evaluated ignoring the direction of error or accounting for its direction. Measures based on the former evaluate forecast precision or accuracy, while measures based on the latter evaluate its bias. Our focus here is on measures of forecast accuracy.

Measures of forecast accuracy can be placed into one of two sets, those that are "scale-dependent" and those that are not (Hyndman and Koehler 2006). Scale-dependent measures should be used with care when making accuracy comparisons across data sets so that different scales which affect the magnitude of these measures are not mis-interpreted as differences in error. The most commonly used scale-dependent summary measures of forecast accuracy are based on the distributions of absolute errors (|E|) or squared errors (E^2), taken over the number of observations (n). These measures include:

Mean Square Error (MSE) = $(\sum E^2) / n$;

Root Mean Square Error (RMSE) = sqrt(MSE);

Mean Absolute Error (MAE) = $(\sum |E|) / n$); and

Median Absolute Error (MEDAE) = median(|E|).

Both MSE and RMSE are integral components in statistical models (e.g., regression). As such, they are natural measures to use in many forecast error evaluations that use regression-based and statistical methods (Alho and Spencer 2005; Pflaumer 1988; Swanson 2008; Swanson and Beck 1994). There is no absolute criterion for a "good" value of any of the scale dependent measures. Moreover, as arithmetic means, the presence of outliers will influence MSE, RMSE, and MAE. As such, they implicitly give greater weight to larger error values. One advantage that

RMSE has over MSE is that its scale is the same as the forecast data. Instead of reporting in terms of the "average" of squared errors, as is the case for MSE, errors reported by the RMSE are representative of the size of an "average" error. MAE is also measured in the same units as the original data, and is usually similar in magnitude to, but slightly smaller than, the RMSE. MEDAE is not influenced by outliers, but this strength is also a weakness in that it does not maximize the use of available information on the errors, a trait it shares with all "robust" measures.

Measures that are not scale-dependent adjust for the population size of the area using a percentage error given by PE = (E / O) * 100. Like the scale dependent measures, a positive value of PE is derived by taking its absolute value (|PE|) or its square (PE^2). These measures include:

Mean Square Percentage Error (MSPE) = $(\sum PE^2) / n$;

Root Mean Square Percentage Error (RMSPE) = $\text{sqrt}((\sum PE^2) / n)$;

Mean Absolute Percentage Error (MAPE) = $(\sum |PE|) / n$; and

Median Absolute Percentage Error (MEDAPE) = median(|PE|).

Because percentage errors are not scale-independent, they are used to compare forecast performance across different data sets. The fact that they assume the existence of a meaningful zero is not a problem in demographic forecasting (as it would be if, for example, one were forecasting temperatures in the Fahrenheit or Celsius scales). However, they have a major disadvantage in that they are infinite or undefined if O = 0 for any observation. Moreover, because the underlying error distributions of these measures have only positive values and no upper bound, percentage errors are highly prone to right-skewed asymmetry in actual practice (Smith and Sincich 1988). This means, for example, that the MAPE is often larger, sometimes substantially larger, than the MEDAPE. The MSPE and RMSPE provides the same properties as the MSE and RMSE, but are expressed as percents.

The symmetrical MAPE (SMAPE) was designed to deal with some of the limitations of the MAPE (Makridakis 1993). Like MAPE, SMAPE is an average of the absolute percent errors but these errors are computed using a denominator representing the average of the forecast and observed values. SMAPE has an upper limit of 200% and offers a well designed range to judge the level of accuracy and should be influenced less by extreme values. It also corrects for the computation asymmetry of the PE. For example, F = 150 and O = 100 yield a PE = 50%, while F = 100 and O =150 yield a PE = 33%. The average of F and O in the denominator of the PE yields 40% in either situation.

Other measures are based on relative errors (Armstrong and Collopy 1992; Hyndman and Koehler 2006; Swanson and Tayman 1995). These measures compare the accuracy from two forecasts, which can be based on different methods and assumptions. They can also compare a forecast from a naïve low cost alternative to one based on a formal forecasting method. Measures based on relative errors are useful for judging the utility of a forecast, or its value in improving the quality of information upon which decisions are based.

3. Mean Absolute Percent Error (MAPE)

Of the preceding measures, MAPE is most commonly used to evaluate cross-sectional forecasts (Ahlburg 1995; Campbell 2002; Hyndman and Koehler 2006; Isserman 1977; Miller 2001; Murdock et al. 1984; Rayer 2007; Sink 1997; Smith 1987; Smith and Sincich 1990, 1992; Smith, Tayman, and Swanson 2001; Tayman, Schaffer, and Carter 1998; Wilson 2007). It is a note of MAPE's ubiquity that it is often found in software packages (e. g., Autobox, ezForecaster, Nostradamus, SAS, and SmartForecast). In addition, MAPE has valuable statistical properties in that it makes use of all observations and has the smallest variability from sample to sample (Levy and Lemeshow 1991). MAPE is also often useful for purposes of reporting, because it is expressed in generic percentage terms that will be understandable to a wide range of users.

MAPE is simple to calculate and easy to understand, which attest to its popularity, but does it meet the criteria for a good measure of error? According to the National Research Council (1980), any summary measure of error should meet five basic criteria — measurement validity, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation. MAPE meets most of these criteria, but its validity is questionable. As noted previously, the distribution of absolute percent errors is often asymmetrical and right skewed. As such, the MAPE is neither a resistant or robust summary measure because a few outliers can dominate it and the MAPE will not be close in value for many distributions (Hoaglin, Mosteller, and Tukey 1983: 28; Huber 1964; Tukey 1970). Therefore, the MAPE can understate forecast accuracy, sometime dramatically. As such, it has tended to reinforce the perception of inaccurate forecasts.

4. Mean Absolute Percent Error Rescaled (MAPE-R)

That MAPE is subject to overstating error because of the presence of extreme outliers has long been known and attempts to constrain the effect of outliers have taken several paths: (1) controlling variables like population size; (2) using a more resistant summary of the distribution like a median or M-estimators; or (3) trimming the tail of the distribution. However, as Swanson, Tayman, and Barr (2000) argued, outliers do inform the improvement of population estimates and forecasts, which is the primary reason they introduced MAPE-R (MAPE-Rescaled). Eliminating outliers removes information and MAPE-R was designed to preserve such information by "normalization" rather than elimination. Two major advantages in using a normalized distribution are that all observations are kept in the analysis, and all measures of central tendency will be approximately the same. Among other things, a normalized distribution suggests that the mean will be as robust and resistant as the median.

To address the effect of a skewed distribution on MAPE, Swanson, Tayman, and Barr (2000) normalized the Absolute Percent Error (APE) distribution using a Box Cox transformation and introduced MAPE-T (MAPE-Transformed) as a summary measure of accuracy for this normalized distribution. The normalized distribution considers the entire data series, but assigns a proportionate amount of influence to each case through normalization, thereby reducing the otherwise disproportionate effect of outliers on a summary measure of error.

To change the shape of a distribution efficiently and objectively and to achieve parity for the observations, Swanson, Tayman, and Barr (2000) used a standardized technique designed to generate a single, nonlinear function to change the shape of the APE distribution. This technique modified the power transformation developed by Box and Cox (1964)[1], defined as:

$y = \left(x^{\lambda} - \lambda\right)/\lambda$ for $x \neq 0$; or

$y = \ln(x)$, for $x = 0$, where

x is the absolute percent error, *y* is the transformed observation, and λ is the power transformation constant.

One determines Lambda (λ) by finding the λ value that maximizes the function:

$$ml(\lambda) = -(n/2) \times \ln\left[(1/n)\sum (y_i - \bar{y})^2\right] + (\lambda - 1) \times \sum LN(x_i), \text{ where}$$

n is the sample size; y is the transformed observation.; $\bar{y}$ is the mean of the transformed observations; x is the original observation.

According to Box and Cox (1964), ml(λ) at a local maximum provides the power transformation (λ) for x that optimizes the *probability* that the transformed distribution will be symmetrical. In other words, finding λ does not guarantee symmetry, but it represents the transformation power most likely to yield a symmetrical distribution. In practice, the maximum value of ml(λ) is found by solving its function for different values of λ between the range of –2 and 2 and identifying the largest resulting Box-Cox value (Draper and Smith 1981: 225).

In preliminary tests, Swanson, Tayman, and Barr (2000) noted that their modified Box-Cox transformation not only compressed very large values, but also increased values greater than one in skewed distributions where λ was relatively small (less than 0.4). This property illustrated how this transformation was more effective in achieving a symmetrical distribution than simpler, non-linear functions that only increased untransformed errors of less than one. Because many estimation errors are greater than one percent, the modified Box-Cox equations not only lowered extremely high values toward the body of the data, but also raised relatively low values. These characteristics minimized skewness and increased symmetry.[2]

The transformed APE distribution has a potential disadvantage: transformation may move the observations into a unit of measurement that is difficult to interpret (Emerson and Stoto 1983: 124). This is not a trivial issue. As mentioned earlier, the National Research Council states that an error measure must have clarity of presentation (National Research Council, 1980). It is easier to think of estimation error in terms of percentages than, for example, log-percentages or square root-percentages. Interpretation may be impeded if the modified Box-Cox transformation is used because it is even less intuitive than simpler transformations, such as the natural log and square root. In addition to reflecting a new unit of measurement, the average error of the transformed distribution may reflect a new scale that further complicates clear understanding and interpretation of error.

Tayman, Swanson, and Barr (1999) suggested that one of two classes of nonlinear functions (quadratic and power) be used to re-express the scale of the transformed observations into the scale of the original observations. Using coefficients from regressions of the APEs on the APE-Ts, they solved for MAPE-R based on the value of MAPE-T. Initially, regression was considered an effective but cumbersome way to re-express MAPE-T into MAPE-R. Testing revealed a more serious problem. When λ approaches zero, regression results become inconsistent. With the closed form expression in mind (as well as the geometric mean), a simple procedure for re-expressing MAPE-T back into the original scale of MAPE was identified by Coleman and Swanson (2007).[3] This re-expression is found by taking the inverse of MAPE-T: [4]

$$\text{MAPE-R} = [(\lambda)(\text{MAPE-T} + 1)]^{1/\lambda}.$$

5. Is MAPE-R Needed?

Swanson, Tayman, and Barr (2000) provide a set of guidelines for determining if MAPE-R is needed. The central issue is the symmetry in the distribution of (APEs). If the distribution of APEs in a given forecast evaluation is symmetrical, then MAPE will appropriately reflect its center of gravity. However, if it is right-skewed, with outliers in the upper tail, then the center of gravity as measured by MAPE is vulnerable to being dominated by these outliers, which suggests that the APEs should be transformed into a more symmetrical distribution. In determining if a set of APEs should be so transformed, Emerson and Stoto (1983: 125) established the following guideline: If the absolute ratio of the highest APE value to the smallest APE value exceeds 20, transformation may be useful; if the ratio is less than 2, then a transformation may not be useful; a ratio between 2 and 20 is indeterminate.

If the Emerson-Stoto guidelines find that a transformation is called for or if the question is indeterminate, Swanson, Tayman, and Barr (2000) suggested using a statistical skewness test to make a final determination in regard to transformation of the APEs. We use the skewness test developed and tested by D'Agostino, Belanger, and D'Agostino Jr. (1990). The null hypothesis tested is that the skewness value = 0, using the 0.10 level of significance. We recommend this significance level rather than more stringent ones (e.g., 0.05 and 0.01) because there is a greater cost in terms of a downwardly biased measure of accuracy in not transforming a potentially skewed distribution.

When the guidelines indicate a potentially useful transformation of APEs to a symmetrical distribution, the transformation is assumed to be successful when the average of the new distribution does not overstate or understate the error level and uses all observations. In this situation, the observations receive nearly equal weights, closer to 1/n, while the resulting average remains intuitively interpretable and clear in its presentation.

6. 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).[5] This data set included 2,481 counties, 79 percent of the national total. For each county, information was collected 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 (the 20 years immediately preceding the launch year), and forecast errors for 10- and 20-year horizons. The launch years included all decennial census years from 1920 to 1990. For this analysis, we selected a 2000 forecast derived from the 1970 to 1990 base period (10-year horizon).[6] Forecast errors were calculated as the percent difference between the population forecasted in 2000 and the population counted in the 2000 decennial census.

Forecasts 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 an average of the forecasts from these five techniques, after excluding the highest and lowest. Simple techniques such as these are frequently used for small-area forecasts and have been found to produce forecasts of total population that are at least as accurate as those produced using more complex or sophisticated techniques (Long 1995; Murdock et al. 1984; Smith and Sincich 1992; Smith, Tayman, and Swanson 2001). An important benefit of these techniques is that they rely on readily available data and can be applied easily to a very large data set. Given the similarity of errors generally found for most forecasting techniques applied to the same geographic regions and time periods, we believe the results reported here are likely to be valid for other techniques and time periods as well.

7. Analysis

We begin by analyzing the entire sample of 2,481 counties (see Table 1). The MAPE of the original APE distribution was 6.21%. The ratio of the highest to lowest APE (5,220) (Max/Min) indicates the need for transformation and the hypothesis of symmetry is rejected (P-value 0.000). Following Tayman, Swanson, and Barr (1999), we use the ratio of the MAPE (MAPE-R) to MEDAPE (median absolute percent error) as an indication of the of bias of the average as a measure of accuracy.[7] For all counties, this ratio is 1.40 suggesting that MAPE understates the average forecast accuracy by 40%.

Table 1 About Here

Figure 1 About Here

The Box-Cox transformation yields a λ value of 0.272 resulting in a MAPE-R 4.42%, which is 29% less than the original MAPE.[8] The MAPE-R to MEDAPE ratio drops to 0.98, indicating that the transformed APE distribution is much less influenced by outlying errors than the original APE distribution. The Box-Cox normalization is successful in that the transformed APE distribution has a skewness coefficient close to zero and the null hypothesis of symmetry is accepted. The Max/Min ratio has decreased substantially to 196; however, if the 10 values of less than 0.5 are excluded that ratio drops to 19.6.

The effect of the transformation is seen in Figure 1, which compares the original APE and transformed APE-T distributions for all counties. The transformation modestly increases the APEs up to the median of the distribution where APE and APR-T are roughly equal. Values above the median are adjusted downward at an increasing rate, which applies the largest adjustments to the most extreme values. As a result, the transformed distribution is no longer influenced by outlying errors and the resulting MAPE-R is smaller than the MAPE.

The usefulness of the Box-Cox normalization and resulting MAPE-R in creating an average measure of forecast accuracy not influenced by outliers is evident when looking at all counties. These results are in line with results reported previously. We now take the analysis a step further by examining the county errors separately for the 48 contiguous states in our sample. We first assess the original APE distribution for the counties in each state to determine whether or not a transformation is warranted. Table 2 provides a summary of the transformation decision and detailed results are found in Table 3.

Eight observations are required to run the skewness test (D'Agostino, Belanger, and D'Agostino Jr. 1990), which eliminates six states (13%) from further analysis. In a skewed distribution with very small samples, the median is a robust measure of central tendency that would not be discarding much information. Seven states (15%) did not require transformation. The hypothesis of symmetry was accepted with P-values greater than 0.10 in these states, and in 5 out of 7 the Max/Min ratio was less than 20. The Max/Min ratio was greater than 20 in New Jersey and South Carolina. In New Jersey, excluding the one value less than 1.5% reduces the ratio to 4.0, while for South Carolina; excluding the 3 values less than 0.3 reduces it to 19.9.

Tables 2 and 3 About Here

Transformations are indicated for 35 states (73%) by both the Max/Min ratio and skewness test. In these states, the Max/Min ratio was greater than 20 and ranges from 39 to over 11,000. The hypothesis of no skewness was rejected in each state at the 0.10 level. Across these states the average of the MAPE/MEDAPE ratios is 1.32, suggesting on average a 32% understatement of forecast accuracy in the original APE distributions. The Box-Cox normalization appears to work well in states whose counties show a skewed APE distribution. The skewness values of the transformed APEs are close to zero (ranging from -0.196 to 0.030) and the hypothesis of no skewness is accepted in each state. The MAPE-R is less than the MAPE for every state, reducing the average error (i.e., increasing the degree of accuracy) by an average of 26%, with a range of 9.2% to 60.6% (last column in Table 3). Moreover, the average of the MAPE-R/MEDAPE ratios is now 0.98, indicating the success of the transformation of achieving measures of average error not influenced by outliers.

To measure the impact of the skewness on the upward bias of the MAPE, we regressed the percent reduction in error against the skewness in the original APE distribution for the 35 states where a transformation was done. A power function using the natural log of both variables best describes this relationship with an adjusted R^2 of 0.508. The elasticity coefficient (0.670) indicates that for every 1% increase in skewness the upward bias of the MAPE increases by approximately 0.7%. Prior to this research, there were an insufficient number of samples to estimate this effect.

A possible concern with this transformation is that it could enhance the data to look "better" than they really are. In fact, the re-expression of the original APE can both increase and decrease the MAPE-T relative to the MAPE. The latter will usually occur with more frequency. In the 35 states, the MAPE-T is smaller the MAPE in 26 and larger in only 9. To illustrate the affect of the transformation for these conditions we examined the error by county in Washington (MAPE = 7.38% and MAPE-T = 4.32) and Pennsylvania (MAPE% = 3.36 and MAPE-T = 8.83).

Figures 2 and 3 About Here

Figure 2 shows the more typical case illustrated by Washington, where a λ of 0.361 results in an upward adjustment of a relatively few small original APEs, modest downward adjustments to errors close to the main body of the data, and an increasing downward adjustment as the APE moves away from the bulk of the observations. In this case, the MAPE-T is portraying an average accuracy that is likely biased downward; therefore the adjustment to MAPE-R increases the average to 6.32%. In Pennsylvania, a different pattern of adjustment emerges (see Figure 3). The smaller λ of 0.110 causes the bulk of the observations to adjust upward. This upward adjustment lessens toward the upper end of the distribution and for the relatively few highest values a substantial downward adjustment occurs. The result is the average of the transformed APE (MAPE-T) is substantially larger than the MAPE and is likely biased upward; therefore the adjustment to MAPE-R decreases that average considerably to 2.04%.

8. Conclusions and Suggestions for Future Research

MAPE is a suitable measure in many instances (Campbell 2002; Rayer 2007). However, it is often based on a positively skewed distribution of APEs, which pulls the average error upward and understates the forecast accuracy of the bulk of the observations. Releasing evaluations that understate accuracy only serves to perpetuate the perception that demographic forecasts are inaccurate. To determine if average error is overstated, we offer a two-step process for evaluating the shape of the original APE distribution. The first step is the Emerson-Stoto test (Emerson and Stoto 1983: 125), which is based on the ratio of the maximum to minimum APE. If the first step fails to suggest that the distribution of APEs does not require transformation, then we advise using a formal hypothesis test of skewness (e.g., D'Agostino, Belanger, and D'Agostino Jr. 1990) to make the final decision.

We showed how these criteria can be used to judge the symmetry of the APE distribution, instead of simply accepting MAPE (or MEDAPE) in a given situation. If these tests indicate that MAPE is not suitable, we suggest using the rescaled MAPE or MAPE-R. We feel that when indicated by the two-step process, MAPE-R represents an improvement over MAPE in evaluating cross-sectional population forecast (and estimate) accuracy. Moreover unlike MEDAPE, MAPE-R preserves information about the structure of error in the presence of the outliers that affect MAPE.

We also demonstrated in this paper how transforming a distribution of APEs normalizes a highly skewed distribution and how to re-express the distribution average (MAPE-T) to provide a fair and useful summary measure of error (MAPE-R) that is robust, resistant, and compliant with the standards set by the National Research Council (1980). We also demonstrated that this procedure works over a wide range of APE distributions and can accommodate situations where original APEs are predominately adjusted either up or down and where the degree of skewness ranges from moderate to extreme. In sum, by using MAPE-T, the information about all of the values in a distribution can be retained while reducing the effect of outliers on the summary measure. By re-expressing MAPE-T into MAPE-R, using the technique described by Coleman and Swanson (2007), MAPE-R is not only easy to calculate and much more consistent in terms of monotonicity, but is also readily understandable and a more accurate summary measure of forecast accuracy than MAPE.

In concluding this empirically-based examination of MAPE-R, we suggest seven areas of future research:

1. Improving the applicability of MAPE-R by developing software that can generally be used in a user friendly computing environment.
2. Exploring the effect on MAPE-R of using an omnibus test of normality that includes both skewness and kurtosis (e.g., Jarque and Bera 1987) instead of a test that just considers skewness.
3. Investigating normality functions with influence curves to address the problems of non-global monotonicity and known instability of the Box-Cox transformation. One possibility is the geometric average (GMAPE), which like MAPE-R is subject neither to the shortcomings observed for MAPE nor to the instability of the Box-Cox transformation.

4. Examining the sensitivity of the λ calibration and how the transformations should be structured. For example, how much would the results by state differ if they were based on a single calibration of all counties rather than a state by state calibration as shown in this paper?
5. Determining if APE-T values can be directly adjusted to match MAPE-R. Currently, the inverse of MAPE-T is used to derive MAPE-R. This means that an analysis of subgroups, such as by size and growth rate, requires computing both MAPE-T and MAPE-R. Having APE-R values could, for example, facilitate the direct computation of MAPE-R for subgroups in a manner similar to using the APE distribution underlying the direct computation of MAPE.
6. Investigating the use of MAPE-R in evaluating the ex post facto accuracy of longitudinal (time series) forecasts, especially in terms of comparing forecast accuracy across series (Hyndman and Koehler 2006). Generally, accuracy decreases as the length of the forecast horizon increases, which suggests that APEs calculated for this type of forecast may be sufficiently skewed to warrant consideration of MAPE-R.
7. Exploring the use of APEs transformed by the Box-Cox function in conjunction with loss functions as an accuracy evaluation tool.[9]

ENDNOTES

1. Swanson; Tayman, and Barr (2000) used λ in the numerator. Box and Cox (1964) used 1.0 in their original development to assure continuity in λ when $\lambda=0$. The difference is immaterial.

2. A potential shortcoming of the Box-Cox transformation is it is not globally monotonic. Individual values may have differential influence on the function. Values near the mean of the transformed distribution have little effect, while extreme outliers may actually *reduce* the MAPE-T. Because the Box-Cox transformation has no associated influence function, it is difficult to determine if and when the Box-Cox will perform this way (Coleman and Swanson, 2007).

3. Coleman and Swanson (2007) find this closed form expression for MAPE-R to be a member of the family of power mean-based accuracy measures. This enables it to be placed in relation to other members of this family, which includes HMAPE (Harmonic Mean Absolute Percent Error), GMAPE (Geometric Mean Absolute Percent Error), and MAPE. Given that MAPE-R was designed to be robust in the face of outliers, it is not surprising to find that it is a valid estimator of the median of the distribution generating the absolute percent errors. Simulation studies suggest that MAPE-R is a far more efficient estimator of this median than MEDAPE

4. If the optimal value of λ found by the Box-Cox procedure is small, between -0.4 and +0.4, the transformed APEs are sufficiently far from the original scale that re-expression is required (Tayman, Swanson, and Barr, 1999).

5. These data were kindly provided by Stefan Rayer, Bureau for Business and Economic Research, University of Florida.

6. This sample of 2,481 counties had average and median sizes in 1990 of 79,100 and 23,400 respectively. Between 1970 and 1990, they grew at an average and median rates of 22% and 14.6%, respectively, with 29% showing population declines during this period.

7. This ratio is used as a descriptive tool to help judge the influence of outliers on the MAPE. We, in effect, are relating two measures of accuracy; one is affected by outliers (MAPE) and the other is not (MEDAPE). In this application, the MEDAPE is a convenient reference point that provides an error measure free of the influence of extreme values.

8. The transformation of very small original APEs (< 0.20%) can result in negative values. These negative values do not affect the resulting MAPE-R and should be set to zero in practice. The transformed values in four (0.2%) counties are slightly negative.

9. Loss functions as a means of forecast and estimation evaluation are explored by, among others, Bryan (2000), Hough and Swanson (2006) and the National Research Council (1980).

REFERENCES

Ahlburg, D. (1992). A Commentary on Error Measures: Error Measures and Choice of a Forecast Method. *International Journal of Forecasting* 8: 99-111.

Ahlburg, D. (1995). Simple versus Complex Models: Evaluation, Accuracy, and Combining. *Mathematical Population Studies* 5: 281-90.

Ahlburg, D., W. Lutz, and J. Vaupel. (1998). Ways to Improve Population Forecasting: What Should Be Done Differently in the Future? *Population and Development Review* 24: 191-198.

Alho, J., S. Jensen, and J. Lassila. (2008). *Uncertain Demographics and Fiscal Sustainability*. Cambridge, England: Cambridge University Press.

Alho, J. and B. Spencer. (1985). Uncertain Population Forecasting. *Journal of the American Statistical Association* 80: 306-314.

Alho, J. and B. Spencer. (2005). Statistical Demography and Forecasting. Dordrecht, The Netherlands: Springer. W. Alonso and P. Starr (Eds.). *The Politics of Numbers*. New York: Russell Sage.

Anderson, M. and S. Fienberg. (2001). *Who Counts? The Politics of Census Taking in Contemporary America.* Revised Paperback Edition. New York, NY: Russell Sage Foundation.

Armstrong, J.S. and F. Collopy. (1992). Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting* 8: 69-80.

Armstrong, J.S. and R. Fildes. (1995). On the Selection of Error Measures for Comparisons among Forecasting Methods. *Journal of Forecasting* 14: 67-71.

Bauer, P. (1981). *Equality, the Third World, and Economic Delusion*. Cambridge, MA: Harvard University Press.

Bell, M. and J. Skinner. (1992). Forecast Accuracy of Australian Subnational Population Projections. *Journal of Population Research* 9 (2): 207-235.

Booth, H. (2004). On the Importance of Being Uncertain: Forecasting Population Futures for Australia. *People and Place* 12(2): 1-12.

Box, G.E.P. and D.R. Cox, (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society, Series* B 26: 211-252.

Brass, W., A. Coale, P. Demeny, D. Heisel, F. Lorimer, A. Romaniuk, and E. Van De Walle. (1968). *The Demography of Tropical Africa. Princeton*. NJ: Princeton University Press.

Bryan, T. 2000. U.S. Census Bureau Population Estimates and Evaluation with Loss Functions. *Statistics in Transition* 4: (4): 537-549.

Bulatao, R. A. (2001). Visible and Invisible Sources of Error in World Population Projections. Paper presented at IUSSP General Conference, Salvador, Brazil, 18–24 August 2001 (http://www.iussp.org/Brazil2001/s40/S40_03_Bulatao.pdf, last accessed January, 2010).

Caldwell, J. and B. Caldwell. (2003). Was there a Neolithic Mortality Crisis? *Journal of Population Research* 20 (2): 153-168.

Campbell, P. (2002). *Evaluating Forecast Error in State Population Projections Using Census 2000 Counts.* Population Division Working Paper Series No. 57. Washington, D.C.: U. S. Census Bureau.

Chamberlain, A. (2006). *Demography in Archaeology*. Cambridge, England: Cambridge University Press.

Choldin, H. (1994). *Looking for the Last Percent: The Controversy over Census Undercounts*. New Brunswick, NJ: Rutgers University Press.

Clements, M. and D. Hendry. (1993). On the Limitations of Comparing Mean Square Forecast Errors. *Journal of Forecasting* 12: 617-637.

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

Coleman, C. and D. Swanson. (2007). On MAPE-R as a Measure of Cross-sectional Estimation and Forecast accuracy. *Journal of Economic and Social Measurement* 32 (4): 219-233.

D'Agostino, R. A. Belanger, and R. D'Agostino, Jr. (1990). A Suggestion for Using Powerful and Informative Tests of Normality. *The American Statistician* 44 (3): 316-321.

de Gans, H. (1999). *Population Forecasting, 1895-1945: The Transition to Modernity*. Dordrecht, The Netherlands: Springer.

Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Thinking*. (Translated by C. Naish) Cambridge, MA: Harvard University Press

Dillman, D., J. Smyth, and L. Christian. (2008). *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 3rd Ed*. Hoboken, NJ: Wiley.

Dorn, H. (1950). Pitfalls in Population Forecasts and Projections. *Journal of the American Statistical Association* 45: 311-34.

Draper, N. and H. Smith. (1981). *Applied Regression Analysis, 2nd Edition*. New York, NY: John Wiley and Sons.

Emerson, J. and M. Stoto. (1983). Transforming Data. pp. 97-128, in: D. Hoaglin, F. Mosteller, and J. Tukey (eds.), Understanding *Robust and Exploratory Data Analysis*. New York, NY: John Wiley.

Hajnal, J. (1955). The Prospects of Population Forecasts. *Journal of the American Statistical Association* 50: 309-22.

Hoaglin, D., F. Mosteller, and J. Tukey. (1983). Introduction to More Refined Estimators pp. 283-296 in: D. Hoaglin, F. Mosteller, and J. Tukey (eds.) *Understanding Robust and Exploratory Data Analysis.* New York: John Wiley.

Hough, G., and D. Swanson. (2006). An Evaluation of the American Community Survey: Results from the Oregon Test Site. *Population Research and Policy Review* 25 (3): 257-273.

Howe, A. (1999). Assessing the Accuracy of Australia's Small-Area Population Estimates. *Journal of Population Research* 16 (1/2): 47-63.

Hull, T. (2007). The Strange History and Problematic Future of the Australian Census. *Journal of Population Research* 24 (1): 1-22.

Huber, P. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35: 73-101.

Hyndman, R. and A. Koehler. (2006). Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting* 22: 679-688.

Isserman, A. (1977). The Accuracy of Population Projections for Subcounty Areas. *Journal of the American Institute for Planners* 43: 247-59.

Isserman, A. and P. Fisher. (1984). Projection, Forecast and Plan: on the Future of Population Forecasting. *Journal of the American Planning Association* 50(2): 208-221.

Jarque, C. and A. Bera. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review* 55 (2): 163–172.

Keilman, N. (1999). How Accurate are the United Nations World Population Projections? *Population and Development Review* 24 (Suppl.):15-41.

Keyfitz, N. (1972). On Future Population. *Journal of the American Statistical Association* 67: 347-363.

Keyfitz, N. (1981). The Limits of Population Forecasting. *Population and Development Review* 7(4): 579-593.

Keyfitz, N. (1982). Can Knowledge Improve Forecasts? *Population and Development Review* 8 (4): 729-751

Keyfitz, N. (1987). The Social and Political Context of Population Forecasting. pp. 235-258 in W. Alonso and P. Starr (Eds.) *The Politics of Numbers*. New York, NY: Russell Sage Foundation.

Lee, L. (2000). *Bad Predictions: 200 Years of the Best Minds Making the Worst Forecasts*. Rochester, MI: Elsewhere Press.

Levy, P. and S. Lemeshow. (1991). *Sampling of Populations: Methods and Applications*. New York: John Wiley.

Long, J. (1995). Complexity, Accuracy, and Utility of Official Population Projections. *Mathematical Population Studies* 5: 203-16.

Makridakis, S. (1993). Accuracy Measures: Theoretical and Practical Concerns. *International Journal of Forecasting* 9: 527-529.

Miller, E. R. (2001). *Evaluation of the 1990 School District Level Population Estimates Based on the Synthetic Ratio Approach*. Population Division Working Paper No. 54. Washington D.C.: US Census Bureau.

Moen, E. (1984). Voodoo Forecasting: Technical, Political and Ethical Issues Regarding the Projection of Local Population Growth. *Population Research and Policy Review* 3(1): 1-25.

Mulder, T. (2002). *Accuracy of the U.S. Census Bureau National Population Projections and Their Respective Components of Change*. Population Division Working Paper No. 50. Washington, D.C.: U.S. Census Bureau. (http://www.census.gov/population/www/documentation/twps0050/twps0050.html, last accessed January 2010).

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

National Research Council. (1980). *Estimating Population and Income for Small Places*. Washington, D.C.: National Academy Press.

National Research Council. (2000). *Beyond Six Billion: Forecasting the World's Population.* Washington, D.C.: National Academy Press.

Pflaumer, P. (1988). Confidence Intervals for Population Projections Based on Monte Carlo Methods. *International Journal of Forecasting* 4:135-142.

Pittenger, D. (1978). The role of Judgment, Assumptions, Techniques, and Confidence Limits in Forecasting Population. *Socio-Economic Planning Sciences* 12(5):271-276.

Preston, S., I. Elo, M. Hill, and I. Rosenwaike. (2003). *The Demography of African Americans 1930-1990.* Dordrecht, The Netherlands: Springer.

Rayer, S. (2007). *Population Forecast Accuracy: Does the Choice of Summary Measure of Error Matter*?" Population Research and Policy Review 26: 163-184.

Rayer, S. (2008). Population Forecast Errors: A Primer for Planners. *Journal of Planning Education and Research*, 27: 417-430.

Romaniuc, A. (1990). Population Projection as Prediction, Simulation, and Prospective Analysis. *Population Bulletin of the United Nations* 29: 16-31.

Romaniuc, A. (1992). Reflection on Population Forecasting: From Predictions to Prospective Analysis. *Canadian Studies in Population* 30 (1): 35-50.

Sanderson, W. (1998). Knowledge Can Improve Forecasts: A Review of Selected Socioeconomic Population Projection Models. *Population and Development Review* 24 (Suppl.):88-117.

Schmitt, R. (1977). *Historical Statistics of Hawaii.* Honolulu, HI: The University of Hawaii Press.

Simon, J. (1996). *The Ultimate Resource, 2nd Edition*. Princeton, NJ: Princeton University Press. (available online as "Ultimate Resource II" at http://www.juliansimon.com/writings/Ultimate_Resource/TCHAR23.txt, last accessed January 2010).

Sink, L. (1997). *Race and Ethnicity Classification Consistency between the Census Bureau and the National Center for Health Statistics*. Population Division Working Paper No. 17. Washington, D.C.: U.S. Census Bureau.

Skerry, P. (2000). *Counting on the Census? Race, Group Identity, and the Evasion of Politics.* Washington, D.C.: The Brookings Institution.

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

Smith, S. and T. Sincich. (1988). Stability over Time in the Distribution of Population Forecast Errors. *Demography* 25: 461-474.

Smith, S. and T. Sincich. (1990). On the Relationship Between Length of Base Period and Population Forecast Errors. *Journal of the American Statistical Association* 85: 367-75.

Smith, S. and T. Sincich. (1991). An Empirical Analysis of the Effect of the Length of Forecast Horizon on Population Forecast Errors. *Demography* 28 (2): 261-274.

Smith, S. and T. Sincich. (1992). Evaluating the Forecast Accuracy and Bias of Alternative Projections for States. *International Journal of Forecasting* 8: 495-508.

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

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

Swanson, D.A. (2008). Measuring Uncertainty in Population Data Generated by the Cohort-Component Method: A Report on Research in Progress. pp. 165-189 in S. Murdock and D. A. Swanson (Eds.) *Applied Demography in the 21st Century*. Dordrecht, The Netherlands: Springer.

Swanson, D.A. and D. Beck. (1994). A New Short-Term County Population Projection Method. *Journal of Economic and Social Measurement* 20: 1-26.

Swanson, D. A., and G. E. Stephan. (2004). Glossary and Demography Timeline. pp. 751 – 786 in J. Siegel and D. A. Swanson (Eds.) *The Methods and Materials of Demography, 2nd Edition*. San Diego, CA: Elsevier Academic Press.

Swanson, D.A. and J. Tayman. (1995). Between A Rock and A Hard Place: The Evaluation of Demographic Forecasts. *Population Research and Policy Review* 14: 233-49.

Swanson, D.A., J. Tayman, and C.F. Barr. (2000). A Note on the Measurement of Accuracy for Subnational Demographic Estimates. *Demography* 37: 193-201.

Sykes, Z. (1969). Some Stochastic Versions of the Matrix Method for Population Dynamics. *Journal of the American Statistical Association* 64: 111-130.

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

Tayman, J. and D.A. Swanson. (1996). On the Utility of Population Forecasts. *Demography* 33: 523-28.

Tayman, J. and D.A. Swanson. (1999). On the Validity of MAPE as a Measure of Population Forecast Accuracy. *Population Research and Policy Review* 18: 299-322.

Tayman, J., D.A. Swanson, and C.F. Barr. (1999). In Search of the Ideal Measure of Accuracy for Subnational Demographic Forecasts. *Population Research and Policy Review* 18: 387-409.

Thornton, R. (1987). *American Indian Holocaust and Survival: A Population History since 1492.* Norman, OK: The University of Oklahoma Press.

Tukey, J. (1970). Exploratory Data Analysis, Vol. I. Reading, MA: Addison-Wesley.

Wilson. T. (2007). The Forecast Accuracy of Australian Bureau of Statistics National Population Projections. *Journal of Population Research* 24 (1): 91-117.

Yokum, J. and J. Armstrong. (1995). Beyond Accuracy: Comparison of Criteria used to Select Forecasting Methods. *International Journal of Forecasting* 11: 591-97.

Zerubavel, E. (2003). *Time Maps: Collective Memory and the Social Shape of the Past.* Chicago, IL: The University of Chicago Press.

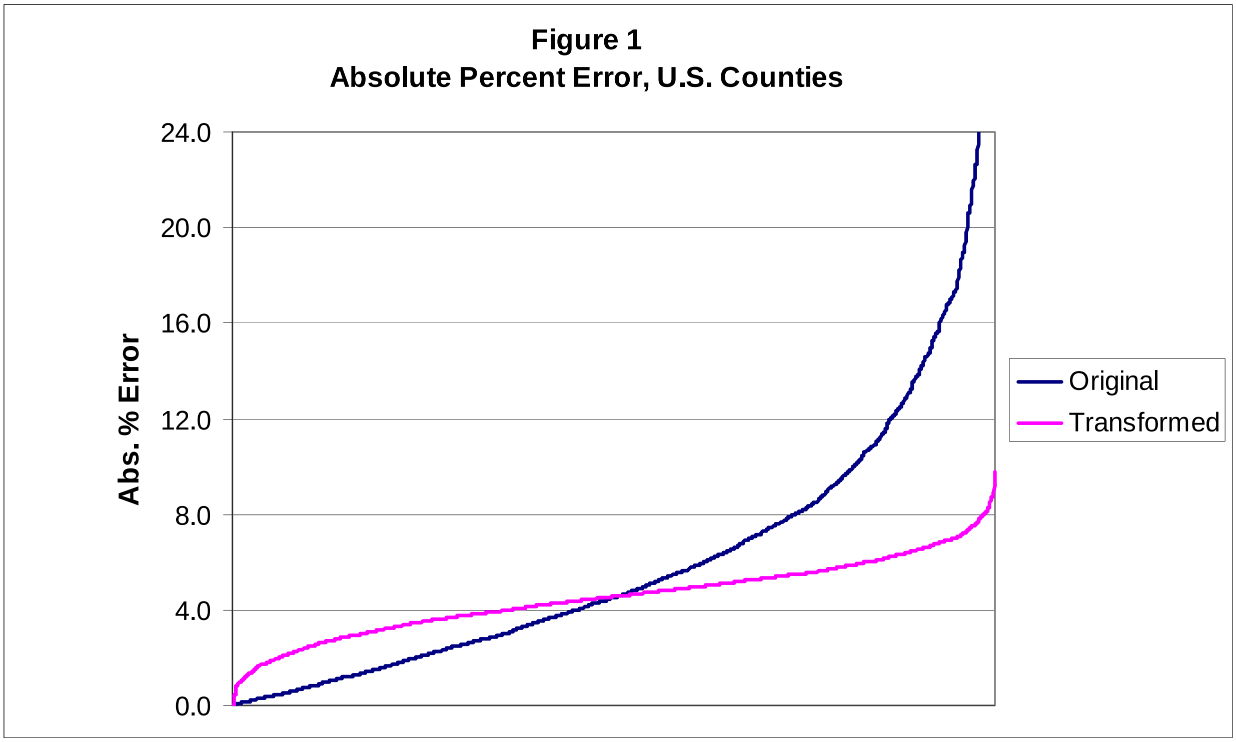
Figure 1
Absolute Percent Error, U.S. Counties
24.0
20.0
16.0
12.0
8.0
4.0
0.0
Abs. % Error
Original
Transformed

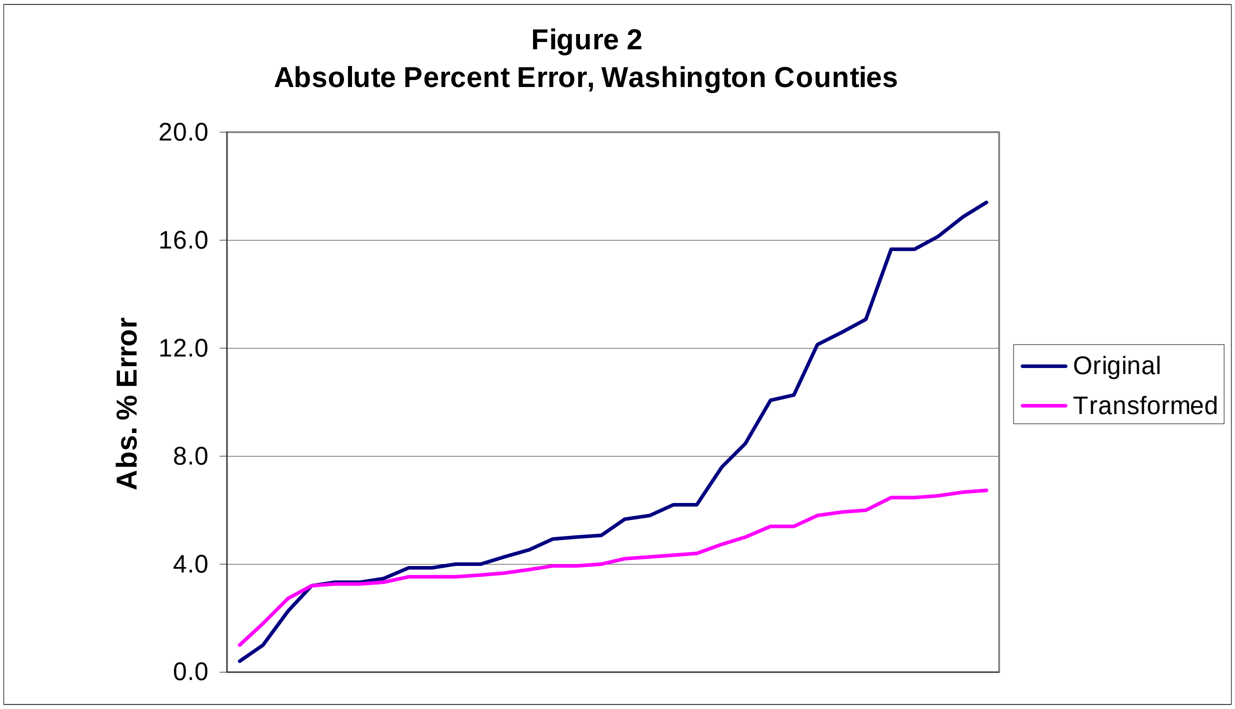
Figure 2
Absolute Percent Error, Washington Counties
20.0
16.0
12.0
8.0
4.0
0.0
Abs. % Error
Original
Transformed

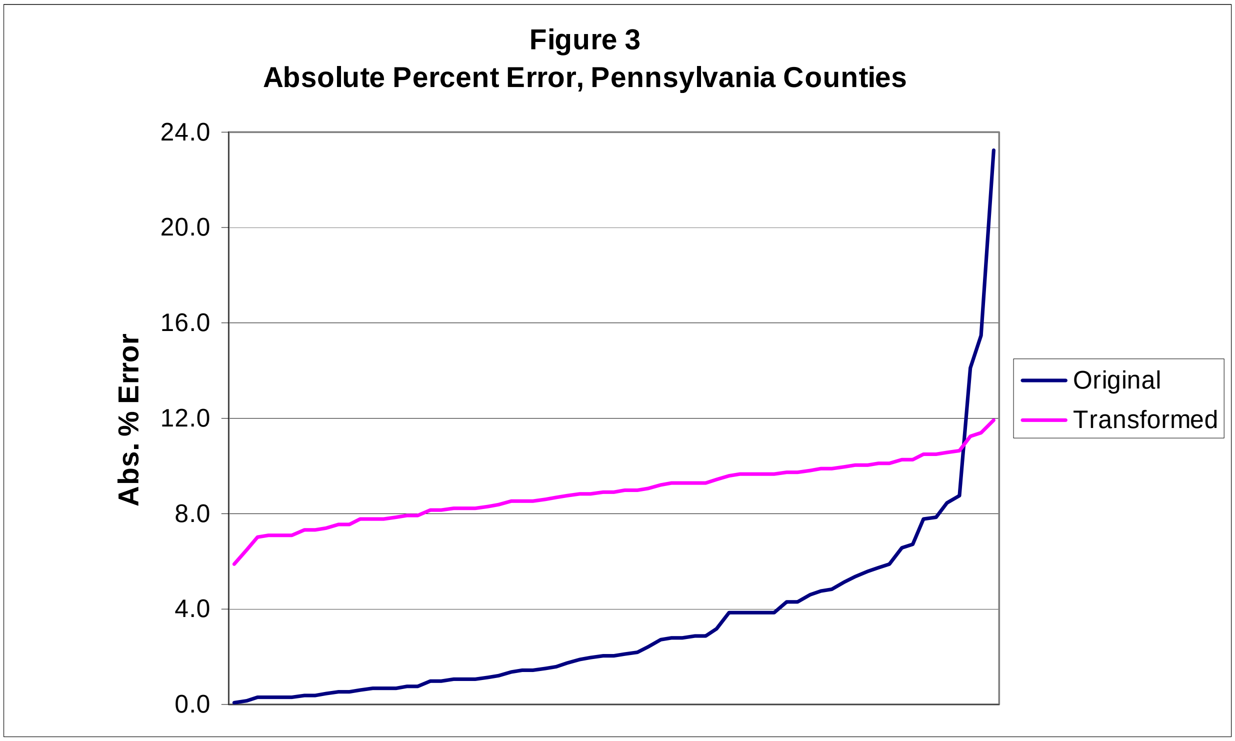
Figure 3
Absolute Percent Error, Pennsylvania Counties
24.0
20.0
16.0
12.0
8.0
4.0
0.0
Abs. % Error
Original
Transformed

Table 1
APE Distribution Statistics, All Counties

	Untransformed	Transformed
Sample Size	2,481	2,481
Lambda	n/a	0.272
Skewness	2.099	-0.017
P-Value[1]	0.000	0.677
Max/Min[2]	5,220	196
MAPE	6.21	n/a
MAPE-R	n/a	4.42

[1] Ho: Skew = 0

[2] Maximum APE / Minimum APE

Table 2
APE Transformation Decision, States

	No.	Percent
Insufficient Sample Size	6	13%
No Transformation Required	7	15%
Transformation Required	35	73%
	48	100%

Table 3
APE Distribution Statistics, States

State	Sample Size	Untransformed MAPE	Untransformed Skew	Untransformed P-Value[1]	Untransformed Max /[2] Min	Transformed Lambda	Transformed MAPE-R	Transformed Skew	Transformed P-Value[1]	% Reduction[3] in Average Error
Insufficient Sample Size										
Delaware	3	8.09	1.150	n/a	17	n/a	n/a	n/a	n/a	n/a
Idaho	4	8.05	1.302	n/a	6	n/a	n/a	n/a	n/a	n/a
Montana	5	11.01	1.453	n/a	7	n/a	n/a	n/a	n/a	n/a
New Mexico	4	7.74	0.894	n/a	4	n/a	n/a	n/a	n/a	n/a
Rhode Island	3	1.99	2.812	n/a	5	n/a	n/a	n/a	n/a	n/a
Wyoming	4	6.77	0.802	n/a	265	n/a	n/a	n/a	n/a	n/a
Transformation Not Required										
Arizona	11	9.96	0.589	0.358	9	n/a	n/a	n/a	n/a	n/a
Connecticut	8	2.80	-0.108	0.884	3	n/a	n/a	n/a	n/a	n/a
Maine	12	8.64	0.051	0.934	8	n/a	n/a	n/a	n/a	n/a
New Hampshire	10	4.30	-0.065	0.922	9	n/a	n/a	n/a	n/a	n/a
New Jersey	21	5.48	0.499	0.300	50	n/a	n/a	n/a	n/a	n/a
South Carolina	18	4.86	-0.067	0.849	60	n/a	n/a	n/a	n/a	n/a
Vermont	14	4.06	0.904	0.125	17	n/a	n/a	n/a	n/a	n/a
Transformation Required										
Alabama	59	4.69	1.228	0.001	157	0.338	3.613	-0.083	0.781	22.9%
Arkansas	62	5.35	1.322	0.000	158	0.314	3.900	-0.093	0.742	27.1%
California	54	6.07	0.836	0.014	255	0.340	4.784	-0.094	0.757	21.2%
Colorado	48	18.89	0.735	0.036	562	0.538	17.150	-0.053	0.872	9.2%
Florida	26	6.92	1.832	0.001	39	0.147	5.104	-0.021	0.956	26.2%
Georgia	101	9.99	1.423	0.000	3,738	0.320	7.244	-0.068	0.760	27.5%
Illinois	102	4.36	1.753	0.000	156	0.194	3.086	-0.029	0.893	29.2%
Indiana	92	4.68	2.148	0.000	11,076	0.407	3.632	-0.083	0.720	22.4%
Iowa	99	4.83	1.625	0.000	77	0.376	4.097	-0.028	0.897	15.1%
Kansas	105	4.66	1.127	0.000	161	0.372	3.783	-0.077	0.729	18.9%
Kentucky	116	6.32	1.887	0.000	3,418	0.335	4.653	-0.052	0.800	26.4%
Louisiana	56	4.98	2.430	0.000	665	0.384	3.939	0.030	0.932	20.9%
Maryland	19	4.01	1.457	0.009	64	0.202	2.785	-0.060	0.899	30.6%
Massachusetts	10	4.78	2.761	0.000	215	0.052	1.884	-0.028	0.962	60.6%
Michigan	83	5.15	2.944	0.000	116	0.090	3.445	-0.009	0.965	33.1%
Minnesota	78	5.78	1.530	0.000	170	0.240	4.364	-0.006	0.974	24.5%
Mississippi	60	5.85	2.025	0.000	5,816	0.303	3.800	-0.081	0.775	35.0%
Missouri	115	5.69	1.453	0.000	92	0.272	4.396	-0.049	0.814	22.7%
Nebraska	87	5.74	1.221	0.000	177	0.405	4.737	-0.054	0.820	17.4%
Nevada	9	11.52	1.418	0.050	39	0.096	6.631	-0.050	0.941	42.5%
New York	56	3.47	1.616	0.000	474	0.304	2.423	-0.096	0.745	30.1%
North Carolina	86	7.42	0.965	0.001	3,702	0.465	6.331	-0.084	0.729	14.7%
North Dakota	31	4.99	0.966	0.026	2,733	0.309	3.162	-0.196	0.619	36.6%
Ohio	88	3.39	3.537	0.000	265	0.133	2.070	-0.009	0.959	38.9%
Oklahoma	9	4.23	2.245	0.003	105	0.206	2.722	-0.018	0.975	35.7%
Oregon	29	6.46	1.414	0.003	58	0.186	4.481	-0.052	0.895	30.6%
Pennsylvania	67	3.36	2.812	0.000	311	0.110	2.039	-0.015	0.951	39.3%
South Dakota	49	7.50	1.121	0.003	94	0.443	6.168	-0.143	0.654	17.8%
Tennessee	94	7.15	0.582	0.022	277	0.500	6.178	-0.145	0.544	13.6%
Texas	232	7.94	1.133	0.000	258	0.279	5.733	-0.074	0.628	27.8%
Utah	25	12.81	0.804	0.082	224	0.469	11.000	-0.066	0.875	14.1%
Virginia	66	7.79	1.943	0.000	48	0.174	6.044	-0.015	0.952	22.4%
Washington	32	7.38	0.789	0.059	42	0.361	6.318	-0.031	0.932	14.5%
West Virginia	55	4.31	1.772	0.000	214	0.226	2.991	-0.048	0.869	30.5%
Wisconsin	64	4.45	0.713	0.021	95	0.519	3.947	-0.107	0.705	11.3%

[1] Ho: Skew = 0
[2] Maximum APE / Minimum APE
[3] (1 - (MAPE / MAPE-R)) *100