MAPE-R: An Empirical Comparison with MAPE

Tom Bryan, Third Wave Research
Madison, Wisconsin, USA
(Tom Bryan [tbryan@thirdwaveresearch.com)

MAPE is simple to calculate and easy to understand, which attest to its popularity, but 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 is easy to present and interpret, but its reliability and validity are questionable. Simple percent differences are affected by the size of the base, which varies widely in small area estimates and forecasts. And the distribution of absolute percent errors is most often asymmetrical and right skewed. Therefore, the mean as a summary measure, can dramatically overstate error.

Although the wisdom of a single-minded focus on accuracy has been questioned (Long, 1995; Pittenger, 1980; Swanson and Tayman, 1995; Tayman and Swanson, 1996), it is widely acknowledged that accuracy is fundamental to estimates and forecasts and their evaluation (Yokum and Armstrong, 1995; Smith, Tayman, and Swanson, 2001; Tayman, Swanson, and Barr, 1999).

Two important properties of a good summary measure are resistance and robustness (Huber, 1964; Tukey, 1970). A measure is resistant if a small subset of the population or sample does not have a disproportionate effect on its value. Resistant measures focus on the main body of the data with little effect from outliers. The median is such a measure; MAPE is not because only a few outliers can dominate it. A related concept is robustness, which is a measure's insensitivity to violations of assumptions underlying it, such as the distributional shape of the observations. Hoaglin, Mosteller, and Tukey (1983: 283) argue that a good summary measure should be close to the true value for many distributions. Because MAPE is a mean, extreme observations have unbounded influence: in the limit, MAPE can become infinite as the result of an infinite error. MAPE is, therefore, not robust. In practice, because MAPE is computed using nonnegative observations, it will increase as its underlying error distribution becomes more asymmetrical and right-skewed.

In a symmetrical distribution, estimates of central tendency are similar or extremely close in value, including the mean, median, mode, and mid-mean (the mean of the central half of the distribution). If the distribution of absolute percent errors (APEs) is symmetric, MAPE *is* a valid measure of location because it neither overstates nor understates the level of error. It also makes use of all observations and has the smallest variability from sample to sample (Levy and Lemeshow, 1991). However, most empirical APE distributions are not symmetric because they are bounded absolutely on the left by zero and are unbounded on the right. Thus, APE distributions are right-skewed, with the degree of skewness determined by the number and values of outliers.

To address the effect of a right-skewed distribution on MAPE, Swanson, Tayman, and Barr (2000) first identified conditions under which the APE distribution should be normalized, using guidelines developed by Emerson and Stoto (1983: 125). They showed how the Box-Cox transformation could be used to effect a normalized the APE distribution and then 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. Because MAPE-T was in effect, in a different scale of measurement than the original APEs, Swanson, Tayman, and Barr (2000) applied an estimate of the inverse of the Box-Cox transformation to rescale MAPE-R back into the original unit of measurement, and termed this final result, "MAPE-R."

Generating MAPE-R: Transformation and Symmetry

To test if a set of APEs should be transformed, Swanson, Tayman, and Barr (2000) used a set of guidelines suggested by Emerson and Stoto (1983: 125): If the (absolute) ratio of the largest value to the smallest value exceeds 20, then transformation is useful. If the ratio is less than two, the transformation may not be useful. A ratio between 2 and 20 is indeterminate. 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, but 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.

To change the shape of a distribution efficiently and objectively and to achieve parity for the observations, Swanson, Tayman, and Barr (2000) used the Box-Cox (Box and Cox, 1964) transformation to render symmetric the APE distribution. The Box-Cox transformation is defined as:

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

(1.a)

$y = \ln(x)$, for $\lambda = 0$, (natural logarithm)

(1.b)

where

x is the absolute percent error: 100 * (estimate/forecast – "truth") / "truth";

y is the transformed observation (which are then averaged to form "MAPE-T," the transformed MAPE); and

λ is the power of transformation.

One determines λ by maximizing the log-likelihood function

$$\ell(\lambda) = -(n/2) \times \ln\left[(1/n)\sum_{i=1}^{n}(y_i - \bar{y})^2\right] + (\lambda - 1) \times \sum_{i=1}^{n} \ln x_i \qquad (1)$$

where

n is the sample size;

y_i is transformed observation i;

$\bar{y}$ is the mean of the transformed observations; and

x_i is original observation i.

The optimal λ minimizes the sum of the squared transformed differences embedded in the log-likelihood (or, equivalently, the likelihood) function and maximizes the probability that the transformed data come from a symmetric normal distribution. The result is the reduction of the absolute value of skewness, $\left|\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^3\right|$, to nearly zero. Values of $\lambda > 1$ eliminate negative skewness, while values of λ, $0 < \lambda < 1$, eliminate positive skewness (Jobson, 1991: 68). It should be noted that the optimal λ does not guarantee symmetry, but, rather, it generates the transformation power most likely to yield a symmetric normal distribution. A possible complication is that the transformed distribution's kurtosis may significantly depart from that of the normal distribution, leading to rejection of tests for normality (Thomas Bryan, personal correspondence).

In their 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, for example, 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 original values 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.

After the distribution of absolute percent errors are transformed, a simple procedure is required to determine the need for re-expression into a measure in the same units as the original APEs. This is done by applying the inverse Box-Cox transformation to MAPE-T:

$$\text{MAPE-R} = [(\lambda)(\text{MAPE-T} + \lambda)]^{1/\lambda}. \qquad (2)$$

and leads to the closed-form expression for MAPE-R:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\text{APE}_i^{\lambda}\right)^{1/\lambda} = M_n^{[\lambda]}(\mathbf{APE}) \qquad (3)$$

where $M_n^{[p]}(\mathbf{a}) = \left(\frac{1}{n}\sum_{i=1}^{n}a_i^p\right)^{1/p}$ is the pth power mean of the positive vector $a = (a_1,\ldots a_n)'$ and APE is the vector of the APE_i.

Test of Utility

The natural extension of this exercise is to develop a full-scale rigorous test of the measure. With the dissemination of U.S. Census 2000 data, demographers were afforded the opportunity to begin testing small-area population estimates that had been developed over the course of the preceding decade. Given the vast improvements in computational power and statistical software, it was disappointing to see the same statistical measures (e.g. quantiles and MAPE) being used to evaluate very large population estimate datasets. These evaluations occasionally undertook crude measures to account for significant outliers, such as by using an arbitrary trimmed mean. Even with these crude and arbitrary measures, the results of these evaluations were very predictable: large percent errors in small pieces of census geography led to very high, biased and unreasonable error measures.

This paper utilizes the MAPE-R as an error measure of block group level population estimates from a private Application Service Provider (ASP) with demographic expertise in the United States. The test involves comparing the private vendor's 2000 block group population estimates with a large sample of the 208,668 Census 2000 block groups in the United States. The exercise will demonstrate the utility of the MAPE_R as a high-quality summary measure of error in small-area population estimates.

Bibliography

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

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: John Wiley.

Hoaglin, D., F. Mosteller, and J. Tukey. 1983. Introduction to more refined estimators. In: D. Hoaglin, F. Mosteller, and J. Tukey (eds.), *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley, 283-296.

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

Jobson, J.D. 1991. *Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design*. New York: Springer-Verlag.

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

National Research Council. 1980. *Estimating Population and Income for Small Places*. Washington, DC: National Academy Press.

Pittenger, D. 1980. Some Problems in Forecasting Population for Government Planning Purposes. *The American Statistician* 34: 135-39.

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

Swanson, D. 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.

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

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.

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

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