Journal of Economic and Social Measurement 32 (2007) 219–233
IOS Press

On MAPE-R as a measure of cross-sectional estimation and forecast accuracy[1]

Charles D. Coleman[a] and David A. Swanson[b]
[a]*Population Division, US Census Bureau, CENHQ Rm. 6H480C, Washington, DC 20233, USA*
E-mail: charles.d.coleman@census.gov
[b]*Department of Sociology and Edward J. Blakely Center for Sustainable Suburban Development, University of California Riverside, Riverside, CA 92521, USA*
E-mail: David.swanson@ucr.edu

Both demographers and economists evaluate the accuracy of their respective forecasts with measures like mean square error, root mean square error, mean absolute percent error, and mean algebraic percent error. However, demographers tend to approach the issue of forecasting very differently than do economists. Two of the distinctive features of the demographic tradition are the use of the cohort-component method (instead of time-series models) and an emphasis on cross-sectional forecasts (instead of forecasts aggregated over time). From the perspective of this demographic tradition, we examine "MAPE-R" (Mean Absolute Percent Error-Rescaled), a recently developed measure of accuracy designed to overcome shortcomings noted in "MAPE" (Mean Absolute Percent Error), a measure commonly used to evaluate the accuracy of population estimates and forecasts. We show that MAPE-R can be calculated simply, thus overcoming the cumbersome calculation procedure used in its introduction and noted as a feature needing correction. We 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(s) generating the absolute percent errors. Simulation studies suggest that MAPE-R is a far more efficient estimator of this median than MEDAPE (Median Absolute Percent Error). Because the Box-Cox transformation on which MAPE-R depends is known to be unstable, we suggest that this represents a line of further research into GMAPE, which, like MAPE-R, is subject neither to the shortcomings observed for MAPE nor to the instability of the Box-Cox transformation. While further lines of research are called for, nothing in our examination of MAPE-R here rules out its use. It also meets the National Research Council's major criteria as a summary measure of accuracy. It is subject to some cautions, but these are no more restrictive than those affecting other accuracy measures, many of which are widely used and have been for some years.

[1]The views expressed in this paper are those of the authors and not necessarily those of the US Census Bureau or the University of California Riverside. An earlier version of this paper was presented at the 2004 meeting of the Southern Demographic Association, Hilton Head, South Carolina., October 14th–16th. We are grateful for comments provided by Tom Bryan, Jeff Tayman, Charlie Barr, Zhenlin Yang, and Tucker McElroy as well as comments and suggestions provided by the editorial staff of the *Journal of Economic and Social Measurement* and anonymous reviewers.

1. Introduction

Both demographers and economists evaluate the accuracy of their respective forecasts with measures like mean square error, root mean square error, mean absolute percent error, and mean algebraic percent error [1,2]. However, demographers tend to approach the issue of forecasting differently than do economists. Two of the distinctive features of the demographic tradition are the use of the cohort-component method (instead of time-series models) and an emphasis on cross-sectional forecasts (instead of forecasts aggregated over time). The distinctly different demographic tradition means, among other things, that turning points in a time-series and the relative accuracy of change versus absolute levels in a time-series are less important to demographers than to economists, even though both use similar (if not the same) software packages and often employ similar (if not the same) language in discussing forecasts and their accuracy. Thus, in this paper, we discuss accuracy not from the perspective of a time-series, but from the perspective of cross-sectional data. However, we argue that the specific measure of accuracy we discuss (MAPE-R) has applicability beyond the demographic perspective.

One characteristic that demographers share with economists is a healthy dose of criticism for their inaccurate estimates and forecasts, with forecasts taking the brunt of the criticism [3–9]. A measure of accuracy commonly used by demographers to evaluate cross-sectional population estimates and forecasts is the *Mean Absolute Percent Error* (MAPE) [1,10–24]. It is a demonstration of MAPE's ubiquity that it is often found in software packages (e.g., SAS, Autobox, Nostradamus, ezForecaster, SmartForecast) that are used not only by demographers, but by economists and others. However, Swanson, Tayman, and Barr [25] argued that MAPE exaggerates error and, as such, tends to contribute to a perception of inaccurate estimates and forecasts.

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) by controlling variables, such as population size; (2) by using a more resistant summary of the distribution like a median; or (3) by trimming the tail of the distribution. However, as Swanson, Tayman, and Barr [25] 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. The normalization is intended to keep all of the observations in the analysis while mitigating the effects of outliers.

This paper derives a closed-form expression for MAPE-R and finds it to be a member of the family of power mean-based accuracy measures. These means, when applied to absolute percent errors, include the familiar MAPE, the geometric mean absolute percent error (GMAPE), and the minimum and maximum absolute percent errors. In addition, the harmonic mean absolute percent error (HMAPE) is a member of this family. An infinite number of these power means exists. An important property that power means possess is that they are in the same units as their arguments. All of the above-mentioned error measures can be interpreted as percentages. MAPE-R has the distinction of having a data-dependent power. The upshot is that MAPE-R is really a known estimator of the underlying median of the APE distribution based on the Box-Cox transformation [26,27].

Given that MAPE-R was designed to be robust in the face of outliers and that it estimates the underlying median of the APE distribution, it is natural to examine its relation to MEDAPE (Median Absolute Percent Error). Simulations of a lognormal distribution and of lognormal mixtures suggest that MAPE-R is a far more efficient estimator of the underlying median than MEDAPE. However, in two of the three scenarios studied, GMAPE, the maximum likelihood estimator of the median of a lognormal distribution [28, p. 220], was even more efficient. This finding suggests that both MAPE-R and GMAPE are useful accuracy measures. Further research is needed to find the circumstances in which MAPE-R is superior to GMAPE and vice versa. This research will also require developing greater understanding of the processes generating estimates errors.

2. Attributes of summary measures

MAPE is simple to calculate and easy to understand, which attest to its popularity, but according to the National Research Council [6] 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 [7, 8,29,30], it is widely acknowledged that accuracy is fundamental to estimates and forecasts and their evaluation [1,30–32].

Two important properties of a good summary measure are resistance and robustness [33,34]. 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 [35, p. 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 asymmetric and right-skewed.

In a symmetric 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 [36]. 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 [25] first identified conditions under which the APE distribution should be normalized, using guidelines developed by Emerson and Stoto [37, p. 125]. They showed how the Box-Cox transformation could be used to effect a normalized 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, essentially, located in a different scale of measurement than the original APEs, Swanson, Tayman, and Barr [25] applied an estimate of the inverse of the Box-Cox transformation to rescale MAPE-T back into the original unit of measurement, and termed this final result, "MAPE-R." They also noted that the rescaling of MAPE-T into MAPE-R was awkward and, as such, an impediment to the use of MAPE-R. This paper provides a much simpler procedure for rescaling MAPE-T into MAPE-R using the exact inverse of the Box-Cox transformation. This, in turn, leads to the closed-form expression for MAPE-R.

3. Generating MAPE-T: Transformation and symmetry

To test if a set of APEs should be transformed to symmetry, Swanson, Tayman, and Barr [25] used a set of guidelines suggested by Emerson and Stoto [37, p. 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, close 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 [25] used the Box-Cox transformation to render symmetric the APE distribution. The Box-Cox transformation is defined as:[1]

$$y = \left(x^{\lambda} - \lambda\right)/\lambda \text{ for } \lambda \neq 0; \text{ or} \tag{1a}$$

$$y = \ln(x), \text{ for } \lambda = 0, \tag{1b}$$

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

y is the transformed observation (all of 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

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

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[2] in the log-likelihood (or, equivalently, likelihood) function and maximizes the likelihood that the transformed data come from a symmetric normal distribution [26, p. 215].[3] The result is the reduction of the absolute value of skewness, $\left|\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^3\right|$, to nearly zero. Values of $\lambda > 1$ eliminate negative skewness, while values of λ, $0 < \lambda < 1$, eliminate positive skewness [38, p. 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.

[1] Box and Cox [26] used 1 in their original development to assure continuity in λ when $\lambda = 0$. As Appendix A below shows, the difference is immaterial to the development of MAPE-R.

[2] That is, λ minimizes $\sum_{i=1}^{n}(z_i - \bar{z})^2$, where $z_i = x_i^{\lambda}\big/\dot{x}^{\lambda}$, where $\dot{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ is the geometric mean of the x_i.

[3] The Box-Cox transformation is really a quasi-maximum likelihood estimator because the left-truncation of the transformed variables at zero eliminates the possibility that they truly come from a normal distribution, which can take on all real values, including non-positive ones [43].

In their preliminary tests, Swanson, Tayman, and Barr [25] 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, for example. This property illustrated how this transformation was more effective in achieving a symmetrical distribution than simpler, nonlinear 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.

The transformed APE distribution has a potential disadvantage: the transformed observations have a different unit of measurement that is difficult to interpret [37, p. 124]. This is because MAPE-T is generated by the APE_i, raised to a power, where i indexes the cross-sectional units, to obtain the transformed APE_i (APET_i). Because MAPE-T is measured in terms of a power of a percentage, intuitive understanding is lost along with the ability to make clear and appropriate interpretations. This is not a trivial issue. As mentioned earlier, the National Research Council [6] states that an error measure must have clarity of presentation.

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.

4. A simple way to Rescale MAPE-T into MAPE-R

As noted by Swanson, Tayman, and Barr [25], their method of rescaling MAPE-T into MAPE-R was cumbersome and they recommended additional research into the possibility that this could be done in a single step. Here we report on just such a procedure.[4] After the distribution of absolute percent errors is transformed, we show that only 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:

$$\text{MAPE-R} = [(\lambda \cdot \text{ MAPE-T} + \lambda)]^{1/\lambda}. \quad (3)$$

As can be seen in a comparison with the procedure introduced by Swanson et al. [25, p. 199], the procedure represented by Eq. (3) is much simpler.

[4] Swanson, Tayman, and Barr [25] used regression to estimate the inverse.

Appendix A shows that Eq. (3) leads to the closed-form expression for MAPE-R:

$$\left(\frac{1}{n}\sum_{i=1}^{n} \text{APE}_i^{\lambda}\right)^{1/\lambda} = M_n^{[\lambda]}\left(\mathbf{APE}\right) \tag{4}$$

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 vector $a = (a_1, \ldots a_n)$ and **APE** is the vector of the APE_i.

5. Characterization of MAPE-R and the power means

The power means are generalizations of the familiar arithmetic mean. Many accuracy measures can be expressed in terms of power means. Important special cases of the power means include the minimum, $M_n^{[-\infty]}(\mathbf{a}) = \min(a_1, \ldots, a_n)$, the harmonic mean, $M_n^{[-1]}(\mathbf{a}) = n\big/\left(\sum_{i=1}^{n} a_i^{-1}\right)$, the geometric mean $M_n^{[0]}(\mathbf{a}) = \left(\prod_{i=1}^{n} a_i\right)^{1/n}$, the arithmetic mean, $M_n^{[1]}(\mathbf{a}) = \frac{1}{n}\sum_{i=1}^{n} a_i$, the root mean square, $M_n^{[2]}(\mathbf{a}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n} a_i^2}$, and the maximum, $M_n^{[\infty]}(\mathbf{a}) = \max(a_1, \ldots, a_n)$. In terms of estimates error, these can be used to define the minimum absolute percent error (MINAPE), the harmonic mean absolute percent error (HMAPE), the geometric mean absolute percent error (GMAPE), the familiar mean absolute percent error (MAPE), the root mean square percent error (RMSPE) and the maximum absolute percent error (MAXAPE).

The power means have the relationship

$$M_n^{[p]}(\mathbf{a}) \leqslant M_n^{[q]}(\mathbf{a}) \tag{5}$$

for $p < q$, with equality obtaining if and only if the a_i are identical [39, p. 85]. This relationship can be applied to estimates error as

$$\begin{aligned}&\text{MINAPE} \leqslant \text{HMAPE} \leqslant \text{GMAPE} \leqslant \text{MAPE-R} \leqslant \text{MAPE} \leqslant \\ &\quad \text{RMSPE} \leqslant \text{MAXAPE}\end{aligned} \tag{6}$$

for $0 < \lambda < 1$, and

$$\text{MINAPE} \leqslant \text{HMAPE} \leqslant \text{GMAPE} \leqslant \text{MAPE} \leqslant \text{MAPE-R} \leqslant \text{MAXAPE} \tag{7}$$

for $\lambda > 1$, with equalities holding if and only if the APE_i are identical.[5]

[5]The relationship between MAPE-R and RMSPE depends on the exact value of λ.

5.1. Perfect estimates

An estimate (or forecast) for an area is perfect if it exactly equals the quantity it estimates. For our purposes, this means at least one of the APE_i is zero, an assumption maintained throughout this section in addition to the assumption that all APE_i are nonnegative. This complicates the use of the power means and the derivation of MAPE-R, as the former are not defined for negative, noninteger powers and the latter always requires positive arguments.[6] These problems are considered separately.

Power Means. Appendix B shows that, for non positive p, the power means are all zero. That is,

$$M_n^{[p]}(\mathbf{a}) = 0 \text{ for all } p \leqslant 0. \tag{8}$$

Let $\mathbf{a}' = (a_1', \ldots, a_m')$ be the vector of positive a_i. (The ordering is unimportant.) Then, Appendix B shows that $M_n^{[p]}(\mathbf{a}) = \left(\frac{m}{n}\right)^{1/p} M_m^{[p]}(\mathbf{a}') > 0$ for positive, finite p, with $M_n^{[\infty]}(\mathbf{a}) = M_m^{[\infty]}(\mathbf{a}') = \max(\mathbf{a}) = \max(\mathbf{a}')$. Inequality (3) continues to hold for positive p and q with equalities obtaining if and only if the a_i' are identical, so that

$$0 = \text{MINAPE} = \text{HMAPE} = \text{GMAPE} < \text{MAPE} \leqslant \text{MAXAPE} \tag{9}$$

with MAPE $=$ MAXAPE if and only if the positive APE_i are identical.

MAPE-R. This measure is complicated by the nonexistence of the Box-Cox transformation when one of the quantities to be transformed is zero per footnote 6. A cure for this is to add a positive quantity, smaller than MINAPE, to each of the APE_i. However, the resulting value of MAPE-R will depend on this quantity.[7] This quantity, however, will not affect the skewness of the augmented APE_i, so that

$$0 = \text{MINAPE} = \text{HMAPE} = \text{GMAPE} < \text{MAPE-R} \leqslant \text{MAPE} \leqslant \text{MAXAPE} \tag{10}$$

for $0 < \lambda < 1$, and

$$0 = \text{MINAPE} = \text{HMAPE} = \text{GMAPE} < \text{MAPE} \leqslant \text{MAPE-R} \leqslant \text{MAXAPE} \tag{11}$$

for $\lambda > 1$, with equalities to the right of GMAPE holding if and only if the positive APE_i are identical.

[6] This can be seen in Footnote # 3, where the geometric mean of the x_i is used as a divisor. This results in division by zero. Alternatively, the log-likelihood Eq. (2) uses logarithms. The logarithm of zero is undefined.

[7] This results from the Box-Cox transformation's lack of scale invariance [44, p. 308].

6. What is MAPE-R estimating?

MAPE-R can be understood by looking at the λth power mean of a randomly generated i.i.d. sample. Let $\mathbf{x} = \{x_i\}_{i=1}^{n}$ be a vector of n i.i.d. random variables. Collins [27, p. 270] shows that the median of the probability density function generating the x_i can be estimated by $M_n^{[\lambda]}(\mathbf{x})$.[8] In fact, this result is true of any sample of independent observations, even when they are not identically distributed. Thus, MAPE-R is an estimator of the median of the distribution (or mixture of distributions) generating the APEs. The statistic $M_n^{[\lambda]}(\mathbf{x})$ is a member of a family of statistics used for the construction of confidence intervals for the median of the distribution generating the x_i [27,40, p. 270]. Unfortunately, we have reason to believe the APEs are non-i.i.d. (in particular, nonidentically distributed, a motivation for the development of MAPE-R), and, consequently, we will not pursue the construction of confidence intervals here. However, it is worthwhile to examine the relationship between MAPE-R and MEDAPE. To do this we run simulations. Before we turn to the results of these simulations, it is useful to note that the MEDAPE for the data found in Swanson, Tayman and Barr [25] data is 4.1902754, while MAPE-R for the same data is 4.0475640, a relative difference of about 3.5 percent (compared to a relative difference of about 20 percent between MAPE-R and MAPE, which is 5.0678707). For the record, GMAPE is 3.6060293.

7. Simulation results and analysis

To further understand the behavior of MAPE-R, we ran three simulations. These simulations used one or more lognormal distributions to generate the APEs. The choice of the lognormal distribution was driven by analysis suggesting that the data of Swanson, Tayman, and Barr [25] are lognormally distributed. Lognormal mixtures generate nonidentically distributed data. Each simulation used 10,000 trials. Three statistics on these distributions were generated and compared: MAPE-R, MEDAPE and GMAPE, the maximum likelihood estimator of the median of the lognormal distribution. The lognormal distribution with parameters μ and σ is the (natural) exponent of a normal distribution with mean μ and standard deviation σ. Its median is simply exp(μ). Throughout the rest of this section, we will use the notation *LN*(μ, σ) to represent this lognormal variable.

[8] The authors would like to thank Zhenlin Yang for pointing out this property. The explanation is that the transformed data are symmetric normal (or approximately so) so that the average of the transformed data is an estimator of their population mean. By symmetry, the median coincides with the mean. Applying the inverse Box-Cox transformation then preserves the quantiles in that a one-to-one mapping exists. Therefore, $M_n^{[\lambda]}(\mathbf{x})$ estimates the median of the x_i. Since the transformation is nonlinear, no such mapping for moments exists.

Table 1
Simulation Results Using Lognormal Errors

	$\hat{\lambda}$	MAPE-R	MEDAPE	GMAPE
Average Error	-0.0008094	0.01335276	0.01738933	0.0123594
Mean Squared Error	0.0149841893	0.0237729376	0.0324758207	0.0204096235
M	14	49	37	83
	(0.7872)	(0.3320)	(0.4654)	(0.0989)
t	-0.66119	8.692457	9.694212	8.683391
	(0.5085)	(<0.0001)	(<0.0001)	(<0.0001)

Note: Two-sided p-values are in parentheses.

Table 2
Simulation Results Using a Lognormal Mixture with Unequal Variances

	$\hat{\lambda}$	MAPE-R	MEDAPE	GMAPE
Average Error	-0.0010612	0.02373321	0.03049062	0.02621021
Mean Squared Error	0.00834085838	0.0493824757	0.0600291516	0.0529925467
M	-93	33	102	25
	(0.0643)	(0.5157)	(0.0424)	(0.6241)
t	-1.16201	10.74087	12.54159	11.45974
	(0.2453)	(<0.0001)	(<0.0001)	(<0.0001)

Note: Two-sided p-values are in parentheses.

Table 3
Simulation Results Using a Lognormal Mixture with Unequal Medians

	$\hat{\lambda}$	MAPE-R	MEDAPE	GMAPE
Average Error	-0.0004538	0.0781274	0.1196514	0.06031024
Mean Squared Error	0.00690147439	0.434510459	0.698482305	0.311951985
M	-13	-64	-39	-2
	(0.8026)	(0.2041)	(0.4413)	(0.9761)
t	-0.54621	11.93586	14.4649	10.86106
	(0.5849)	(<0.0001)	(<0.0001)	(<0.0001)

Note: Two-sided p-values are in parentheses.

Simulation 1 uses the distribution *LN*(0, 1). Simulation 2 uses a lognormal mixture with equal μ and unequal σ. In this simulation, each APE has equal probabilities (50%) of being drawn from either *LN*(0, 1) or *LN*(0, 2). Simulation 3 uses a lognormal mixture with unequal μ and equal σ. Each APE has a 50% probability of being drawn from *LN*(0, 1) or *LN*(1, 1). In each case, the population λ equals 0. Tables 1–3 report the results of Simulations 1-3, respectively. They show the average error (AE), mean squared error (MSE), McNemar's M statistic for median-bias and the t statistic for mean-bias, along with the p-values for the bias statistics. Median-bias is actually an older concept than mean-bias, the now ordinary definition of bias. If an estimator is median-unbiased, then it has equal probabilities of being above or below the parameter. A desirable property of any estimator is that it be both mean- and median-unbiased.

All of the simulations show that MEDAPE is the worst performing estimator of the median of distribution generating the APEs. In every case, MEDAPE has the highest

AE and MSE. While every estimator is mean-biased, MEDAPE is median-biased in Simulation 2, unlike MAPE-R and GMAPE. As expected, in Simulation 1, GMAPE is the best performing estimator with the lowest AE and MSE. However, GMAPE is median-biased, while MAPE-R is not. In Simulation 2, MAPE-R is the best estimator with lowest AE and MSE. GMAPE wins Simulation 3 with the lowest AE and MSE. Note that Simulation 3's MSEs are an order of magnitude larger than those of the other simulations.

The simulations show that both GMAPE and MAPE-R are viable estimators of the underlying median APE. Their superior performance vis- à-vis MEDAPE results from their use of information in the whole sample, while MEDAPE use only the single observation in the middle. GMAPE and MAPE-R show evidence of differential superiority: one works better in some situations than others. More research, both theoretical and empirical, is needed on the distributions of estimates errors to better inform the choice between them.

It is important to note that, for fixed $\lambda = 0$, GMAPE $=$ MAPE-R. So, if estimates errors can be shown to be lognormally distributed, then GMAPE is not only an intuitively understandable accuracy measure, but also one that can be interpreted in terms of the error distribution itself.

8. Discussion

In this paper we have provided a simple means of calculating MAPE-R, thus removing a major barrier to its use noted by Swanson, Tayman, and Barr [25]. We also find that MAPE-R is not only a member of the family of power means, but that it estimates the same quantity that MEDAPE estimates. When λ is fixed at 0, MAPE-R coincides with GMAPE, the maximum likelihood estimator of the median of a lognormal distribution.[9] We also note that MAPE-R represents a possible paradox in that it is a power mean-based accuracy measure only in the formal sense, as it estimates the median, a quantile and not a power mean, of the APE distribution. This occurs because the data determine the power λ needed to induce symmetry in the APE-T and the transformation is inverted back to the original scale, an operation that preserves quantiles, but not moments.

Andrews [41] found evidence of instability in the Box-Cox transformation. This, in turn, suggests an examination of the relationship between MAPE-R and GMAPE, a measure that is not based on the Box-Cox transformation, but, importantly, one that satisfies nearly all of the National Research Council's [6] criteria for an accuracy statistic. Moreover, GMAPE is insensitive to the direction of error. GMAPE, however, has a cost in that it cannot deal with perfect estimates. The reason for

[9] Relative advantages and disadvantages of accuracy measures are discussed in a demographic context by Swanson, Tayman, and Barr [25], Tayman and Swanson [9], and Tayman, Swanson, and Barr [32].

this is that its influence function diverges at zero. Except for this instance, and the fact that its interpretation is less straightforward than either MAPE or MAPE-R [6], it is a reliable and useful accuracy measure. Among related avenues of research would be to estimate absolute percentage error distributions as lognormal mixtures to determine if multiple error generation processes are present.

While further lines of research are called for, nothing in our examination of MAPE-R here suggests that it should not be used. It is now less cumbersome to calculate and, as such, more accessible. It also meets the National Research Council's major criteria as a summary measure of accuracy. Like other measures, it is subject to some cautions, but these are no more restrictive than those affecting other measures, many of which are widely used and have been for some years [11,15,20,24,42]. We conclude by noting that with the appropriate considerations that MAPE-R may also be of interest to those who approach forecasting from the economic perspective rather than the demographic one.

Appendix A: Exact derivation of MAPE-R

MAPE-R is constructed by averaging the generalized Box-Cox transformations of the APE_i, then applying the inverse transformation to this average. The generalized Box-Cox transformation is

$$y_i = f(x_i, c) = \begin{cases} \frac{x_i^\lambda - c}{\lambda} & \lambda \neq 0 \\ \log x_i & \lambda = 0 \end{cases}$$

for all c and for all $x_i > 0$. The base of the logarithm is unimportant. The inverse transformation is defined as

$$x_i = f^{-1}(y_i, c) = \begin{cases} (\lambda y_i + c)^{1/\lambda} & \lambda \neq 0 \\ \exp y_i & \lambda = 0 \end{cases}.$$

Here, $f(\text{APE}_i)$ is set equal to APE-T_i, the transformed APE_i. MAPE-T, the mean of the APE-T_i, is defined for $\lambda^* \neq 0$, where λ^* is the optimal value of the Box-Cox transformation parameter, as

$$\text{MAPE} - \text{T} = \frac{\frac{1}{n}\sum_{i=1}^{n} (\text{APE}_i)^{\lambda^*} - c}{\lambda^*}$$

MAPE-R is thus

$$\text{MAPE} - \text{R} = f^{-1}(\text{MAPE} - \text{T}, c) = (\lambda^* \text{MAPE} - \text{T} + c)^{1/\lambda^*}$$

$$= \left(\lambda^* \frac{\frac{1}{n}\sum_{i=1}^{n} (\text{APE}_i)^{\lambda^*} - c}{\lambda^*} + c \right)^{1/\lambda^*} = \left(\frac{1}{n} \sum_{i=1}^{n} (\text{APE}_i)^{\lambda^*} \right)^{1/\lambda^*} = M_n^{[\lambda^*]}(\mathbf{APE}).$$

When $\lambda^* = 0$, either the Box-Cox transformation or the 0^{th} power mean can be used to obtain GMAPE.

This derivation also shows that MAPE-R is the quasi-arithmetic non-symmetric mean of APE using f [39, p. 75].

Appendix B: Power means with zero arguments

This has several cases, which are considered in turn. All the a_i are assumed nonnegative, while at least one a_i is assumed to be zero. We will use the vector of positive a_i, $\mathbf{a}' = (a'_1, \ldots, a'_m)$. (The ordering is unimportant.)

Positive p:

Case 1a: $p = \infty$:

$$M_n^{[\infty]}(\mathbf{a}) = \max(\mathbf{a}', \mathbf{0}_{n-m}) = M_m^{[\infty]}(\mathbf{a}') = \max(\mathbf{a}'),$$

where $\mathbf{0}_{n-m}$ is the vector of zeroes in *a*. Thus, the inclusion of a zero a_i has no effect on the definition of this power mean.

Case 1b: $0 < p < \infty$:

$$M_n^{[p]}(\mathbf{a}) = \left(\tfrac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p} = \left(\tfrac{1}{n}\sum_{j=1}^m \left(a'_j\right)^p\right)^{1/p}$$

$$= \left(\tfrac{m}{n}\tfrac{1}{m}\sum_{j=1}^m \left(a'_j\right)^p\right)^{1/p} = \left(\frac{m}{n}\right)^{1/p} M_m^{[p]}(\mathbf{a}')$$

Nonpositive p:

Case 2a: $p = 0$:

$$M_n^{[0]}(\mathbf{a}) = \left(\prod_{i=1}^n a_i\right)^{1/n} = 0$$

because at least one of the $a_i = 0$.

Case 2b: $p = -\infty$:

$$M_n^{[-\infty]}(\mathbf{a}) = \min(a_1, \ldots, a_n) = 0$$

because at least one of the $a_i = 0$.

Case 2c: $-\infty < p < 0$:

This case requires the most care, as calculations are performed directly on the expression for $M_n^{[p]}(\mathbf{a})$.

$$M_n^{[p]}(\mathbf{a}) = \left(\tfrac{1}{n}\sum_{i=1}^n a_i^p\right)^{1/p} = \infty^{1/p} = 0.$$

The second equality results from raising zero to a negative power. The last equality follows from raising infinity to a negative power.

Cases 2a-c combine to show that $M_n^{[p]}(\mathbf{a}) = 0$ for all $p \leqslant 0$.

References

[1] S. Smith, J. Tayman and D.A. Swanson, *State and Local Population Projections: Methodology and Analysis*, New York: Kluwer Academic/Plenum Publishers, 2001.
[2] S. Makridakas, S. Wheelwright and R. Hyndman, Forecasting: Methods and Applications. New York, NY: Wiley, 1998.
[3] H. Dorn, Pitfalls in Population Forecasts and Projections, *Journal of the American Statistical Association* **45** (1950), 311–334.
[4] M.V. George, S. Smith, D.A. Swanson and J. Tayman, Population Projections, in: *The Methods and Materials of Demography*, (2nd Edition), New York: Elsevier Academic Press, 2004, pp. 561–601.
[5] J. Hajnal, The Prospects of Population Forecasts, *Journal of the American Statistical Association* **50** (1955), 309–322.
[6] National Research Council, *Estimating Population and Income for Small Places*. Washington, DC: National Academy Press, 1980
[7] D.A. Swanson and J. Tayman, Between A Rock and A Hard Place: The Evaluation of Demographic Forecasts, *Population Research and Policy Review* **14** (1995), 233–249.
[8] J. Tayman and D.A. Swanson, On the Utility of Population Forecasts, *Demography* **33** (1996), 523–528.
[9] J. Tayman and D.A. Swanson, On the Validity of MAPE as a Measure of Population Forecast Accuracy, *Population Research and Policy Review* **18** (1999), 299–322.
[10] D. Ahlburg, Simple Versus Complex Models: Evaluation, Accuracy, and Combining, *Mathematical Population Studies* **5** (1995), 281–290.
[11] T. Bryan, Population Estimates, in: *The Methods and Materials of Demography*, (2nd Edition), New York: Elsevier Academic Press, 2004, pp. 523–560.
[12] A. Isserman, The Accuracy of Population Projections for Subcounty Areas, *Journal of the American Institute for Planners* **43** (1977), 247–259.
[13] E.R. Miller, 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, 2001.
[14] S. Murdock, L. Leistritz, R. Hamm, S. Hwang and B. Parpia, An Assessment of the Accuracy of a Regional Economic-Demographic Projection Model, *Demography* **21** (1984), 383–404.
[15] M. Shahidullah and M. Flotow, Criteria for selecting a suitable method for producing post-2000 county population estimates: A case study of population estimates in Illinois, *Population Research and Policy Review* **24**(3) (2005), 215–229.
[16] L. Sink, 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.: US Census Bureau, 1997.
[17] S. Smith, Tests of Accuracy and Bias for County Population Projections, *Journal of the American Statistical Association* **82** (1987), 991–1003.
[18] S. Smith and T. Sincich, On the Relationship Between Length of Base Period and Population Forecast Errors, *Journal of the American Statistical Association* **85** (1990), 367–375.
[19] S. Smith and T. Sincich, Evaluating the Forecast Accuracy and Bias of Alternative Projections for States, *International Journal of Forecasting* **8** (1992), 495–508.
[20] D.A. Swanson, Advancing Methodological Knowledge within State and Local Demography: a Case Study, *Population Research and Policy Review* **23**(2) (2004), 379–398.
[21] D.A. Swanson and D. Beck, A new short-term county population projection method, *Journal of Economic and Social Measurement* **20**(1) (1995), 25–50.
[22] D.A. Swanson, J. Tayman and D. Beck, On the utility of lagged ratio-correlation as a short-term county population projection method: a case study of Washington State, *Journal of Economic and Social Measurement* **21**(1) (1995), 1–16.
[23] J. Tayman, E. Schafer and L. Carter, 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** (1998), 1–20.

[24] J. Tayman, S. Smith and S. Rayer, Stability over Time in the Distribution of County Population Forecast Errors. Paper presented at the annual Conference of the Southern Demographic Association. University of Mississippi. November 4th, 2005.
[25] D.A. Swanson, J. Tayman and C.F. Barr, A Note on the Measurement of Accuracy for Subnational Demographic Estimates, *Demography* **37** (2000), 193–201.
[26] G.E.P. Box and D.R. Cox, An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B* **26** (1964), 211–252.
[27] S. Collins, Prediction techniques for Box-Cox Regression Models, *Journal of Business and Economic Statistics* **9** (1991), 267–277.
[28] N.L. Johnson, S. Kotz and N. Balakrishnanan, *Continuous Univariate Distributions*, Volume 1. New York: John Wiley, 1994.
[29] J. Long, Complexity, Accuracy, and Utility of Official Population Projections, *Mathematical Population Studies* **5** (1995), 203–216.
[30] D. Pittenger, Some Problems in Forecasting Population for Government Planning Purposes, *The American Statistician* **34** (1980), 135–139.
[31] J. Yokum and J. Armstrong, Beyond Accuracy: Comparison of Criteria Used to Select Forecasting Methods, *International Journal of Forecasting* **11** (1995), 591–597.
[32] J. Tayman, D.A. Swanson and C.F. Barr, In Search of the Ideal Measure of Accuracy for Subnational Demographic Forecasts, *Population Research and Policy Review* **18** (1999), 387–409.
[33] P. Huber, Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics* **35** (1964), 73–101.
[34] J. Tukey, *Exploratory Data Analysis, Vol. I.* Reading, MA: Addison-Wesley, 1970.
[35] D. Hoaglin, F. Mosteller and J. Tukey, Introduction to more refined estimators, in: *Understanding Robust and Exploratory Data Analysis*, D. Hoaglin, F. Mosteller and J. Tukey, eds, New York: John Wiley, 1983, pp. 283–296.
[36] P. Levy and S. Lemeshow, *Sampling of Populations: Methods and Applications*, New York: John Wiley, 1991.
[37] J. Emerson and M. Stoto, Transforming Data, in: *Understanding Robust and Exploratory Data Analysis,* D. Hoaglin, F. Mosteller and J. Tukey, eds, New York: John Wiley, 1983, pp. 97–128.
[38] J.D. Jobson, *Applied multivariate data analysis*, New York: Springer-Verlag, 1991.
[39] D.S. Mitrinović and P.M. Vasić, *Analytical Inequalities*, New York: Springer Verlag, 1970.
[40] Z. Yang, Median estimation through a regression transformation, *The Canadian Journal of Statistics* **30** (2002), 235–242.
[41] D.F. Andrews, A Note on the Selection of Data Transformations, *Biometrika* **58** (1971), 249–254.
[42] G. Hough and D.A. Swanson, Toward an assessment of continuous measurement: A comparison of returns with 1990 census returns for the Portland test site, *Journal of Economic and Social Measurement* **24**(3–4) (1998), 295–308.
[43] C. Coleman, A Fast, High-Precision Implementation of the Univariate One-Parameter Box-Cox Transformation Using the Golden Section Search in SAS/IML®. *Proceedings of the 17th Northeast SAS Users Group* (*NESUG*) *Conference,* http://www.nesug.org/html/Proceedings/nesug04/an/an12.pdf, 2004.
[44] J. Schlesselman, Power families: A note on the Box and Cox Transformation, *Journal of the Royal Statistical Society, Series B* **33** (1971), 307–311.