International Journal of Forecasting 32 (2016) 669–679

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

A new metric of absolute percentage error for intermittent demand forecasts

Sungil Kim [a], Heeyoung Kim [b,*]

[a] Data Analytics Lab, R&D Center, Samsung SDS, Seoul, Republic of Korea

[b] Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

ARTICLE INFO

Keywords:
Accuracy measure
Forecast evaluation
Intermittent demand
MAPE

ABSTRACT

The mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale-independency and interpretability. However, MAPE has the significant disadvantage that it produces infinite or undefined values for zero or close-to-zero actual values. In order to address this issue in MAPE, we propose a new measure of forecast accuracy called the *mean arctangent absolute percentage error* (*MAAPE*). MAAPE has been developed through looking at MAPE from a different angle. In essence, MAAPE is a *slope as an angle*, while MAPE is a *slope as a ratio*, considering a triangle with adjacent and opposite sides that are equal to an actual value and the difference between the actual and forecast values, respectively. MAAPE inherently preserves the philosophy of MAPE, overcoming the problem of division by zero by using bounded influences for outliers in a fundamental manner through considering the ratio as an angle instead of a slope. The theoretical properties of MAAPE are investigated, and the practical advantages are demonstrated using both simulated and real-life data.

1. Introduction

The mean absolute percentage error (MAPE) is one of the most popular measures of the forecast accuracy. It is recommended in most textbooks (e.g., Bowerman, O'Connell, & Koehler, 2004; Hanke & Reitsch, 1995), and was used as the primary measure in the M-competition (Makridakis et al., 1982). MAPE is the average of absolute percentage errors (APE). Let A_t and F_t denote the actual and forecast values at data point t, respectively. Then, MAPE is defined as:

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{A_t - F_t}{A_t}\right|, \qquad (1.1)$$

where N is the number of data points. To be more rigorous, Eq. (1.1) should be multiplied by 100, but this is omitted in this paper for ease of presentation without loss of generality. MAPE is scale-independent and easy to interpret, which makes it popular with industry practitioners (Byrne, 2012).

However, MAPE has a significant disadvantage: it produces infinite or undefined values when the actual values are zero or close to zero, which is a common occurrence in some fields. If the actual values are very small (usually less than one), MAPE yields extremely large percentage errors (outliers), while zero actual values result in infinite MAPEs. In practice, data with numerous zero values are observed in various areas, such as retailing, biology, and finance, among others. For the area of retailing, Fig. 1 (Makridakis, Wheelwright, & Hyndman, 1998) illustrates typical intermittent sales data. Many zero sales occur dur-

* Corresponding author. Tel.: +82 42 350 3125; fax: +82 42 350 3110.
E-mail address: heeyoungkim@kaist.ac.kr (H. Kim).

http://dx.doi.org/10.1016/j.ijforecast.2015.12.003

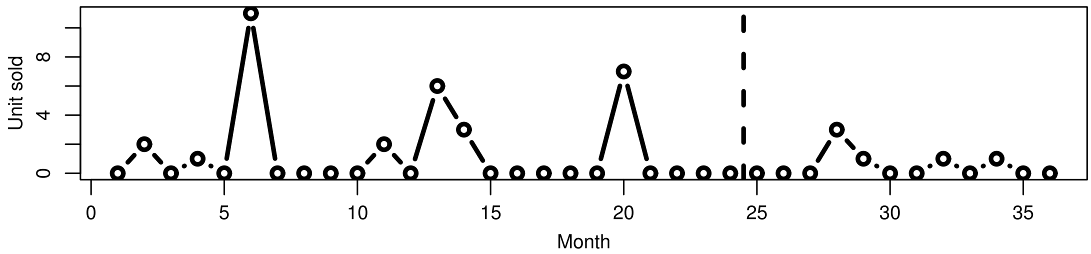


Fig. 1. Three years of monthly sales of a lubricant product sold in large containers. Data source: 'Product C' from Makridakis et al. (1998, Ch. 1). The vertical dashed line indicates the end of the data used for fitting and the start of the data used for out-of-sample forecasting.

ing the time periods considered, and this leads to infinite or undefined MAPEs.

There have been attempts to resolve this problem by excluding outliers that have actual values less of than one or APE values greater than the MAPE plus three standard deviations (Makridakis, 1993). However, this approach is only an arbitrary adjustment, and leads to another question, namely how the outliers can be removed. Moreover, the exclusion of outliers might distort the information provided, particularly when the data involve numerous small actual values. Several alternative measures have been proposed to address this issue. The symmetric mean absolute percentage error (sMAPE), proposed by Makridakis (1993), is a modified MAPE in which the divisor is half of the sum of the actual and forecast values. Another measure, the mean absolute scaled error (MASE), was proposed by Hyndman and Koehler (2006). The MASE is obtained by scaling the forecast error based on the in-sample mean absolute error using the naïve (random walk) forecast method, and can overcome the problem of the MAPE generating infinite or undefined values. Similarly, Kolassa and Schütz (2007) proposed that the mean absolute error be scaled by the in-sample mean of the series (MAE/Mean ratio) in order to overcome the problem of division by zero.

While these alternative measures resolve the MAPE's issue with outliers, the original MAPE remains the preferred method of business forecasters and practitioners, due to both its popularity in the forecasting literature and its intuitive interpretation as an *absolute percentage error*. Therefore, this paper proposes an alternative measure that has the same interpretation as an *absolute percentage error*, but can overcome the MAPE's disadvantage of generating infinite values for zero actual values.

Even though this paper focuses on MAPE, it is worth reviewing the other accuracy measures used in the literature as well. In general, accuracy measures can be split into two groups: scale-dependent measures and scale-independent measures. As the group names indicate, the scale-dependent measures are measures for which the scale depends on the scale of the data. The mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and median absolute error (MdAE) all belong to this category. These measures are useful when comparing different forecasting methods that are applied to data with the same scale, but should not be used when comparing forecasts for series that are on different scales (Chatfield, 1988; Fildes & Makridakis, 1988). In that situation, scale-independent measures are more appropriate. Being scale-independent has been considered to be a key characteristic for a good measure (Makridakis, 1993). The aforementioned MAPE, sMAPE, MASE, and the MAE/Mean ratio are examples of scale-independent measures.

There have been various attempts in the literature to make scale-dependent measures scale-independent by dividing the forecast error by the error obtained from a benchmark forecasting method (e.g., a random walk). The resulting measure is referred to as a relative error. The mean relative absolute error (MRAE), median relative absolute error (MdRAE), and geometric mean relative absolute error (GMRAE) all belong to this category. Even though Armstrong and Collopy (1992) recommended the use of relative absolute errors, particularly the GMRAE and MdRAE, these measures have the issue of potentially involving division by zero. In order to overcome this difficulty, Armstrong and Collopy (1992) recommended that extreme values be trimmed; however, this increases both the complexity and the arbitrariness of the calculation, as the amount of trimming must be specified.

Relative measures are another type of scale-independent measure. Relative measures are similar to relative errors, except that relative measures are based on the values of measures instead of errors. For example, the relative MSE (RelMSE) is given by the MSE divided by MSE_b, where MSE_b denotes the MSE from a benchmark method. Similar relative measures can be defined using RMSE, MAE, MdAE, MAPE, and so on. A log-transformed RelMSE, i.e., log(RelMSE), has also been proposed, in order to impose symmetrical penalties on the errors (Thompson, 1990). When the benchmark method is a random walk and the forecasts are all one-step forecasts, the relative RMSE is Theil's U statistic (Theil, 1966, Ch. 2), which is one of the most popular relative measures. However, Theil's U statistic has the disadvantages that its interpretation is difficult and outliers can easily distort the comparisons because it does not have an upper bound (Makridakis & Hibon, 1979). In general, relative measures can be highly problematic when the divisor is zero. For a more in-depth review of other accuracy measures, refer to Hyndman and Koehler (2006), who provide an extensive discussion of various measures of forecast accuracy, and Hyndman (2006), particularly for measures for intermittent demand.

The remainder of this paper is organized as follows. In Section 2, MAPE is investigated from a different angle, with a new measure called MAAPE being proposed as a result. The behavior and theoretical properties of the proposed measure are then investigated in Section 3. In Section 4, we further explore the bias aspect of MAAPE in comparison

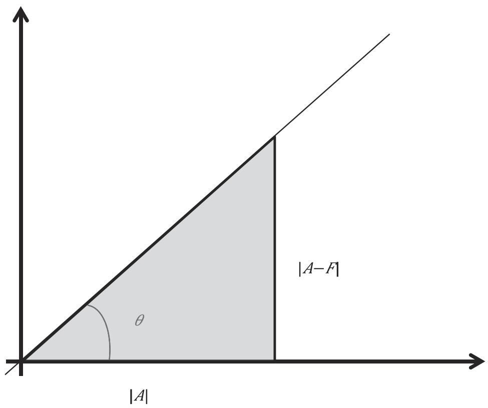


Fig. 2. Conceptual justification of AAPE: AAPE corresponds to the angle θ, while APE corresponds to the slope as a ratio = $\tan(\theta) = \left|\frac{A-F}{A}\right|$, where A and F are the actual and forecast values, respectively.

with MAPE. Then, in Section 5, MAAPE is applied to both simulated and real-life data, and compared with other measures. Finally, we conclude in Section 6.

2. MAPE from a different angle: slope as a ratio vs. slope as an angle

We investigate MAPE from a different angle and propose a new measure of the forecast accuracy. Recall that MAPE is the average of the absolute percentage error (APE). We consider a triangle with adjacent and opposite sides that are equal to $|A|$ and $|A-F|$, respectively, where A and F are the actual and forecast values, respectively, as depicted in Fig. 2. In principle, APE can be viewed as the slope of the hypotenuse. Clearly, the slope can be measured either as a *ratio* of $|A - F|$ to $|A|$, ranging from zero to infinity; or, alternatively, as an *angle*, varying from 0 to 90°. Given that the *slope as a ratio* is the APE, the *slope as an angle* has the potential to be a useful measure of the forecast accuracy, as we propose in this paper. Note that, for the slope, the ratio is the tangent of the angle. Then, the angle θ can be expressed using $|A|$ and $|A - F|$ as follows:

$$\theta = \arctan(\text{ratio}) = \arctan\left(\left|\frac{A-F}{A}\right|\right), \tag{2.1}$$

where 'arctan' is the arctangent (or inverse tangent) function.

Using Eq. (2.1), we propose a new measure, called the mean arctangent absolute percentage error (MAAPE), as follows:

$$\text{MAAPE} = \frac{1}{N}\sum_{t=1}^{N}(\text{AAPE}_t) \quad \text{for } t = 1, \ldots, N, \tag{2.2}$$

where

$$\text{AAPE}_t = \arctan\left(\left|\frac{A_t - F_t}{A_t}\right|\right).$$

Recall that the function $\arctan x$ is defined for all real values from negative infinity to infinity, and $\lim_{x\to\infty} \tan^{-1} x = \pi/2$. With a slight manipulation of notations, for the range $[0, \infty]$ of APE, the corresponding range of AAPE is $[0, \frac{\pi}{2}]$.

3. Properties of MAAPE

This section compares MAPE and MAAPE, in order to investigate the properties of MAAPE. Recall that APE and AAPE are defined by components of MAPE and MAAPE, as in Eqs. (1.1) and (2.2), respectively. Without loss of generality, we therefore compare APE and AAPE.

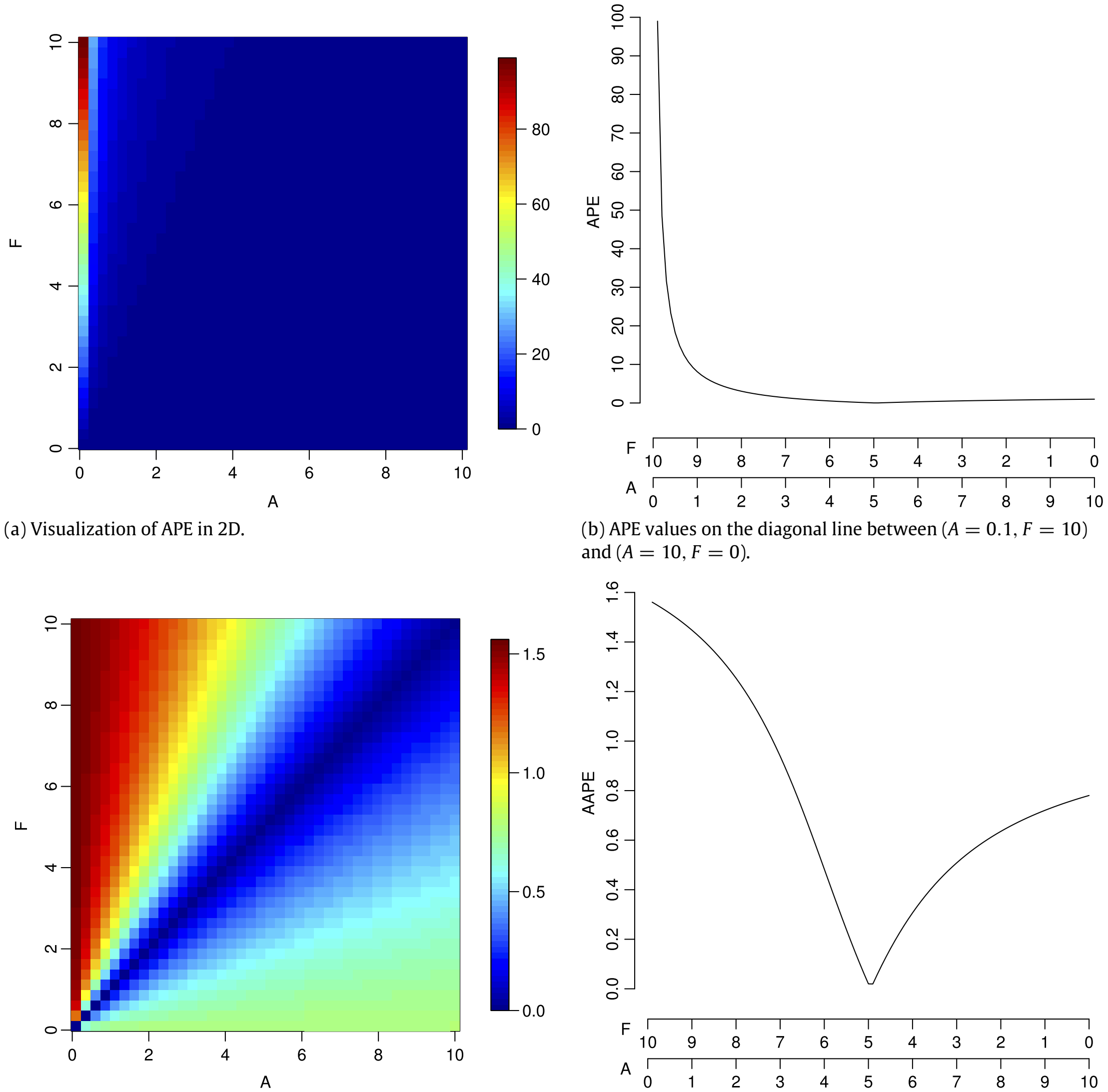
Fig. 3 provides visualizations of APE and AAPE in the upper and lower rows, respectively, with actual (A) and forecast (F) values that vary from 0.1 to 10 in increments of 0.1. In the left column, the values of each measure are presented in a color map, varying from blue (low values) to red (high values). The actual and forecast values are on the x- and y-axes, respectively. For example, in Fig. 3(a), the upper-left corner presents APE values for small actual values and large forecast values, while the lower-right corner presents APE values for large actual values and small forecast values. As expected, the APE values in the upper-left corner are much larger than those in other regions. In the right column, the values of each measure on the diagonal line of the corresponding figure in the left column (from upper-left to lower-right) are plotted. On the x-axis in Fig. 3(b), both actual (A) and forecast (F) values are presented; for simplicity, the x-axis can be regarded as F/A. Fig. 3(a) and (b) clearly illustrate the drawbacks of MAPE: it provides extremely large values when the actual values are small. In contrast, it can be seen clearly in Fig. 3(c) and (d) that AAPE does not go to infinity even with close-to-zero actual values, which is a significant advantage of MAAPE over MAPE. It is evident from a comparison of Fig. 3(c) and (d) with Fig. 3(a) and (b) that AAPE is less sensitive to small actual values than APE.

Due to the value of APE going to infinity for close-to-zero actual values, a detailed comparison of the behaviors of APE and AAPE in Fig. 3 is not easy. Therefore, in order to compare the behaviors of APE and AAPE in more detail, we focus on the actual and forecast values from 1 to 10 instead. Fig. 4 visualizes the APE and AAPE values on a diagonal line similar to those in Fig. 3(b) and (d). As is shown in Fig. 4(a), the value of APE is significantly larger when the forecast value is greater than the actual value (i.e., a positive error) than for the opposite case (i.e., a negative error). MAPE has been criticized because it places significantly heavier penalties on positive errors than on negative errors (Makridakis, 1993). Using a log-transformed version of MAPE (Swanson, Tayman, & Barr, 2000) or the log of the ratio of the predicted to actual values (Tofallis, 2015) can overcome the problem that the percentage errors are not distributed symmetrically, but they retain the problem of division by zero.

Note that APE is a concave-up function for $F > A$, meaning that the unbalanced penalty becomes more severe as $\frac{F}{A}$ increases, which results in APE becoming rapidly larger as the actual value approaches zero. In contrast, AAPE has a more balanced penalty than APE, as Fig. 4(b) shows. This is because AAPE is bounded by $[0, \frac{\pi}{2}]$ and is a concave-down function for $F > 2A$, as Theorem 1 proves.

Theorem 1. *For $0 < A < F$, AAPE is a concave-down function for $F > 2A$ and a concave-up function for $A < F < 2A$, with $F = 2A$ being an inflection point.*

The proof of Theorem 1 is provided in Appendix A.



(a) Visualization of APE in 2D.

(b) APE values on the diagonal line between ($A = 0.1, F = 10$) and ($A = 10, F = 0$).

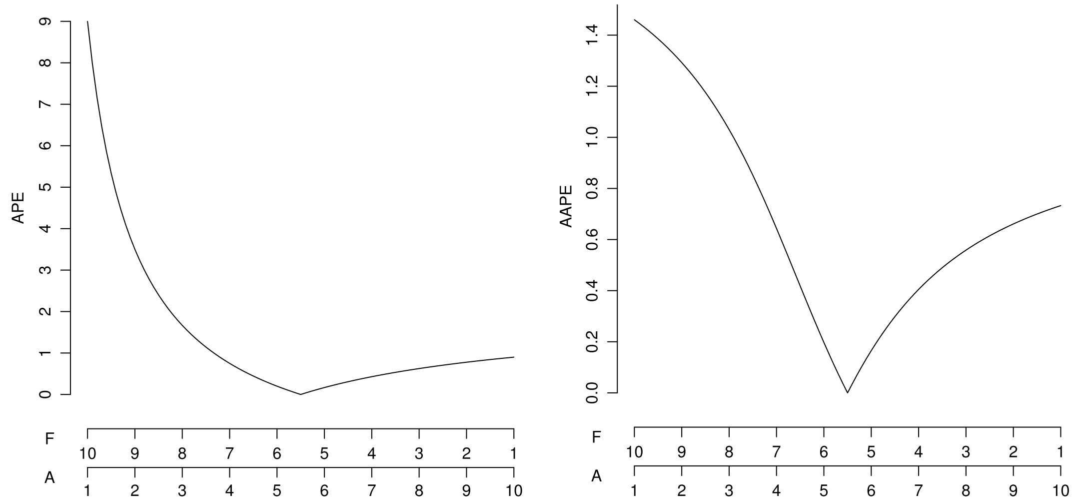
(c) Visualization of AAPE in 2D.

(d) AAPE values on the diagonal line between ($A = 0.1, F = 10$) and ($A = 10, F = 0$).

Fig. 3. Visualization of APE and AAPE on $[0.1, 10] \times [0.1, 10]$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The behavior of AAPE is investigated further in Fig. 5. The x-axis represents APE, and three different lines are compared on the y-axis: $y = \arctan(x)$, $y = x$, and $y = \pi/2 - 1/x$. As the figure shows, $\arctan(x)$ behaves like $y = x$ for small values of x. As x increases, it becomes a nonlinear function, eventually converging to $\pi/2$. This implies that AAPE converges to $\pi/2$ for sufficiently large forecast errors. For a forecast error of more than 316% ($x = 3.16$), the difference between $\arctan(x)$ and $\pi/2 - 1/x$ becomes less than 0.01. In many cases, more than 300% of forecast errors might be regarded as outliers. The convergence of AAPE for large forecast errors plays a part in limiting the influence of outliers, which often distort the calculation of the overall forecast accuracy. This property of AAPE helps make the MAAPE robust against outliers. Therefore, MAAPE can be particularly useful if there are extremely large forecast errors as a result of mistaken or incorrect measurements.

However, if the extremely large forecast errors are considered as genuine variations that might have some important business implications, rather than being due to mistaken or incorrect measurements, MAAPE would not be appropriate. The simple example in Table 1 illustrates the circumstances under which MAAPE is or is not recommended. Consider the two sequences of actual demands illustrated in Table 1: the two sequences have the same values for each period, except for the second period, in which Actual 1 has the value 100, while Actual 2 has the value 100,000. We assume that the same sequence of forecasts is obtained for each sequence of actual demands as in Table 1: the demand for each period is forecast correctly, except for the second period. As a result, MAAPE is calculated to be 0.29 for Actual 1 and 0.31 for Actual 2;



(a) APE. (b) AAPE.

Fig. 4. Visualization of APE and AAPE on [1, 10] × [1, 10].

Table 1
A simple example that illustrates the circumstances under which MAAPE is or is not recommended.

	Period 1	Period 2	Period 3	Period 4	Period 5	MAAPE
Actual $1_{(AAPE)}$	$1_{(0.00)}$	$100_{(1.46)}$	$1_{(0.00)}$	$5_{(0.00)}$	$1_{(0.00)}$	0.29
Actual $2_{(AAPE)}$	$1_{(0.00)}$	$100,000_{(1.57)}$	$1_{(0.00)}$	$5_{(0.00)}$	$1_{(0.00)}$	0.31
Forecast	1	10	1	5	1	

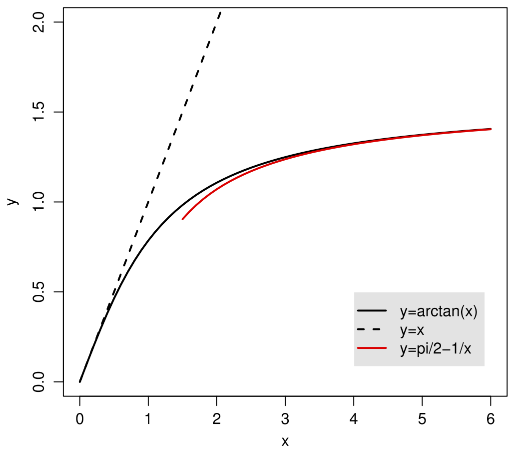


Fig. 5. For small values of x, $\arctan(x)$ varies linearly with x, with its variation becoming nonlinear with increasing values of x; it eventually approaches $\pi/2$.

thus, the values of MAAPE for Actual 1 and Actual 2 do not differ significantly, even though their actual values at Period 2 are significantly different. Therefore, if the actual values 100 and 100,000 are both considered to be mistaken or incorrect measurements, MAAPE is robust against these outliers, and thus MAAPE is recommended for use. In contrast, if the larger actual value of 100,000 compared with the actual value of 100 has important business implications, MAAPE is not sensitive to these large errors, and therefore is not recommended for use. Another limitation of MAAPE is that if the actual value is zero, then the corresponding AAPE value is always $\pi/2$, regardless of the forecast value. Syntetos (2001) noted that sMAPE has a similar limitation: the symmetric absolute percentage error is always equal to two for a zero actual value, regardless of the forecast that is used.

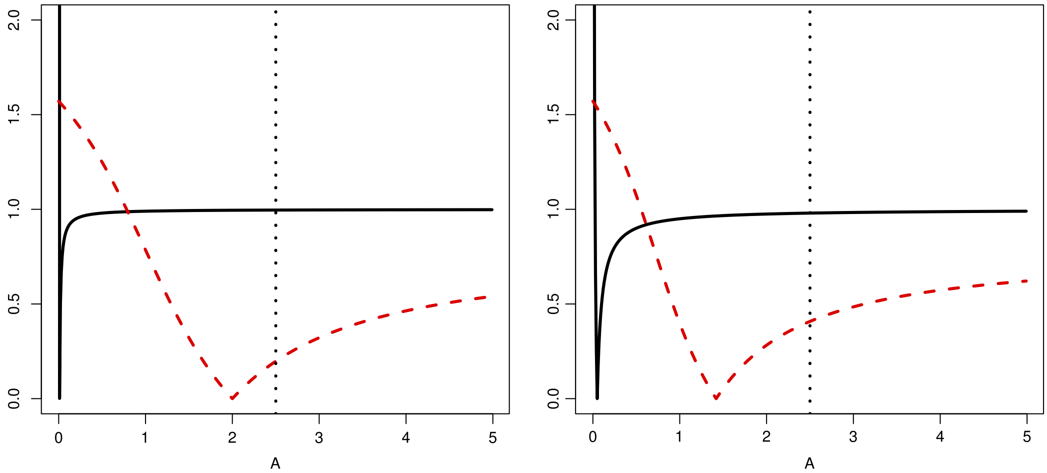
4. Optimal point predictions under MAAPE

This section compares MAPE and MAAPE in terms of one-point predictions, and demonstrates that MAAPE is less biased than MAPE. We assume that a forecaster makes a one-point forecast value of F by solving one of the three objective problems: (1) minimize the total expected loss, (2) equate the expected loss above and below the point prediction, or (3) minimize the maximum possible loss. These three strategies were considered in order to address the bias issue of MAPE in McKenzie (2011). The forecast value obtained by solving the above objective problem is called the *optimal point prediction*. The optimal point predictions corresponding to these objective functions are as follows:

$$F_L^{(1)} = \operatorname*{argmin}_F \left\{ \sum_{A=A_L}^{A_H} L(A,F)P(A) \right\}, \tag{4.1}$$

$$F_L^{(2)} = \operatorname*{argmin}_F \max \left\{ \sum_{A=A_L}^{F-1} L(A,F)P(A), \sum_{A=F}^{A_H} L(A,F)P(A) \right\}, \tag{4.2}$$

$$F_L^{(3)} = \operatorname*{argmin}_F \max \{L(A,F)\}, \tag{4.3}$$



(a) $APE(A, F_{APE}^{(1)})$ and $AAPE(A, F_{AAPE}^{(1)})$.

(b) $APE(A, F_{APE}^{(2)})$ and $AAPE(A, F_{AAPE}^{(2)})$.

Fig. 6. Comparison between $APE(A, F_{APE}^{(i)})$ (black solid line) and $AAPE(A, F_{AAPE}^{(i)})$ (red dashed line) for (a) $i = 1$ and (b) $i = 2$, for A that follows a discrete uniform distribution. The vertical dotted line indicates the locations of both the mean and the median of the actual demand distribution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $L(A, F)$ is a loss function (error measure) that quantifies the deviation of a forecast value (F) from its actual value (A), A_L and A_H are the lower and upper bounds of A, respectively, and $P(A)$ is the probability mass function of the actual value distribution; that is, we assume that A is a non-negative integer random variable that follows a certain discrete probability distribution.

We investigate $F_{AAPE}^{(1)}$, $F_{AAPE}^{(2)}$, and $F_{AAPE}^{(3)}$ when $L(A, F)$ is AAPE, and compare these with $F_{APE}^{(1)}$, $F_{APE}^{(2)}$, and $F_{APE}^{(3)}$, respectively, which are the optimal point predictions under APE. For the third objective function in Eq. (4.3), $F_L^{(3)}$ does not depend on $P(A)$. In this case, we can prove that $F_{AAPE}^{(3)}$ and $F_{APE}^{(3)}$ are identical as in Proposition 1.

Proposition 1. *$F_{APE}^{(3)}$ and $F_{AAPE}^{(3)}$ are identical.*

The proof of Proposition 1 is in Appendix B.

The other two optimal point predictions depend on the specification of $P(A)$, for which we consider both the discrete uniform distribution and the negative binomial distribution. For each distribution, we use a simulation to compare $F_{APE}^{(i)}$ and $F_{AAPE}^{(i)}$, where $i = 1, 2$, in reference to the mean and median of the actual value distribution. For the discrete uniform distribution, A_L and A_H are set to 0 and 5, respectively. For the negative binomial distribution, the following probability mass function was considered ($A_L = 0$ and $A_H = \infty$):

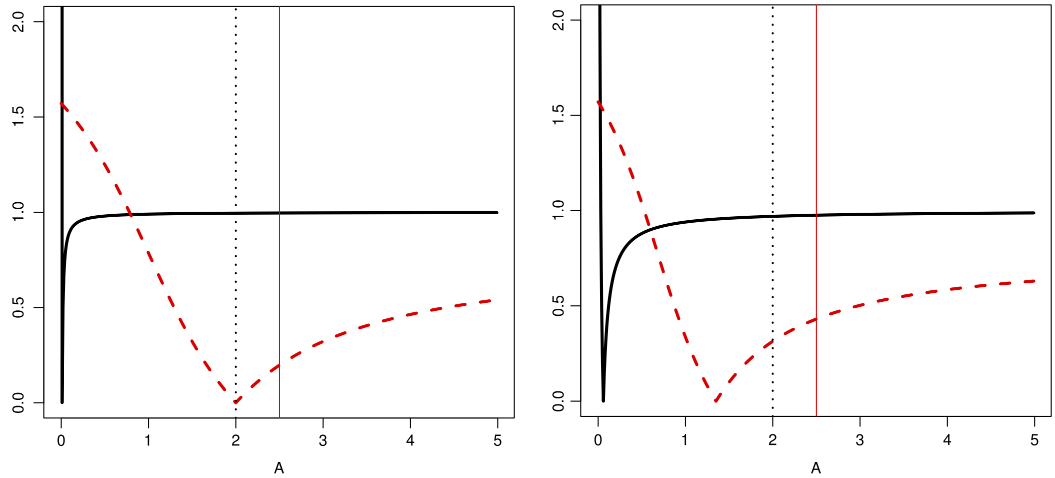
$$P(A = x|r, p) = \binom{r + x - 1}{x} p^r (1 - p)^x, \quad x = 0, 1, 2, \ldots, \tag{4.4}$$

with the parameters $r = 3.58$ and $p = 0.59$, so that we have $E[A] = 2.5$, for ease of comparison with the uniform distribution case. The negative binomial distribution was investigated by Syntetos, Lengu, and Babai (2013), and found to be a good representation of intermittent demand patterns.

Then, under the assumption that the actual values are realized from each distribution, the optimal point predictions under APE and AAPE are compared in Fig. 6 for the discrete uniform distribution and in Fig. 7 for the negative binomial distribution. Fig. 6(a) (or Fig. 7(a)) and Fig. 6(b) (or Fig. 7(b)) correspond to the objectives in Eqs. (4.1) and (4.2), respectively. In the figures, the black solid line and the red dashed line represent the values of APE and AAPE, respectively, and the minimum values of the two lines indicate the locations of $F_{APE}^{(i)}$ and $F_{AAPE}^{(i)}$. Because the APE is undefined for $A = 0$, we added 0.01 to $A = 0$ when calculating the APE.

In the case of the discrete uniform distribution in Fig. 6, the black vertical dotted line indicates both the mean and the median of the actual demand distribution. Fig. 6 demonstrates that both APE and AAPE yield optimal forecasts that are less than the mean (or the median) for the two objectives; however, the optimal forecast under AAPE is closer to the mean (or the median) than that under APE. In the case of the negative binomial distribution in Fig. 7, the solid red and dotted black vertical lines indicate the mean and the median of the actual demand distribution, respectively. Fig. 7 shows the behaviors of APE and AAPE to be similar to those in Fig. 6: the optimal forecasts under APE and AAPE are both less than the mean for both objectives; however, the optimal forecast under AAPE is less biased than that under APE. When considering the median, $F_{APE}^{(i)}$ is less than the median for both $i = 1, 2$, while $F_{AAPE}^{(2)}$ is less than the median and $F_{AAPE}^{(1)}$ is equal to the median.

The simulation results demonstrate that, while both APE and AAPE yield downward-biased forecasts, those under AAPE are significantly less biased than those under APE. This is due to the loss function of APE having significantly heavier penalties when F is larger than A; thus, the optimal point prediction becomes significantly smaller than the mean, in order to minimize the overall penalties. The loss function of AAPE removes the imbalance



(a) $APE(A, F_{APE}^{(1)})$ and $AAPE(A, F_{AAPE}^{(1)})$. (b) $APE(A, F_{APE}^{(2)})$ and $AAPE(A, F_{AAPE}^{(2)})$.

Fig. 7. Comparison between $APE(A, F_{APE}^{(i)})$ (black solid line) and $AAPE(A, F_{AAPE}^{(i)})$ (red dashed line) for (a) $i = 1$ and (b) $i = 2$, for A that follows a negative binomial distribution ($r = 3.58$, $p = 0.59$). The solid red and dotted black vertical lines indicate the locations of the mean and median of the actual demand distribution, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in the penalties for positive and negative errors. With respect to the median, the optimal point prediction under APE is less than the median, while the optimal point prediction under AAPE is less than or equal to the median, depending on the objective function and the distribution of A. Teunter and Duncan (2009) and Morlidge et al. (2015) noted that the point prediction under MAE is optimized on the median for the objective in Eq. (4.1); thus, the optimal point prediction under MAE for the objective in Eq. (4.1) is always zero when at least 50% of the observations are zero. The optimal point prediction under MAAPE will also become zero if there are numerous zero observations, because the optimal point prediction under MAAPE is less than or equal to the median.

5. Applications

The performance of MAAPE was evaluated using two real datasets and a simulated dataset. In Sections 5.1 and 5.2, the advantages of MAAPE over MAPE, sMAPE, MASE, and the MAE/Mean ratio are illustrated using the monthly sales dataset of a lubricant product and a dataset of weekly retail sales, respectively. Section 5.3 compares the performances of MAAPE and MAPE further using a simulated example.

5.1. Example 1: monthly sales data of a lubricant product

Recall the monthly sales data of a lubricant product from Section 1. Fig. 8 depicts the actual data from the 25th month to the 36th month ($A_{25} \sim A_{36}$), along with two different forecasts: $F_{1,25} \sim F_{1,36}$ and $F_{2,25} \sim F_{2,36}$. F_{1t} was computed by simply using the mean of the actual data from the 1st to 24th months. That is, $F_{1,k} = \frac{1}{24}\sum_{t=1}^{24} A_t$ for $k = 25, \ldots, 36$. F_{2t} represents a more accurate forecast, except for one outlier in the 28th month. Because the forecast method itself is not the focus of this paper, we

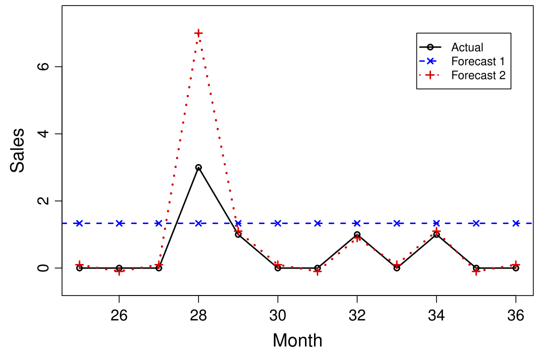


Fig. 8. Actual data and two different forecasts from the 25th to 36th months.

Table 2
The results of MAPE, MAAPE, sMAPE, MASE, and the MAE/Mean ratio for the two different forecasts.

	Excluding $A_t = 0$	Including $A_t = 0$
$MAPE_1$	0.39	∞
$MAPE_2$	0.41	∞
$MAAPE_1$	0.37	1.17
$MAAPE_2$	0.31	1.15
$sMAPE_1$	0.41	1.47
$sMAPE_2$	0.27	1.42
$MASE_1$	0.26	0.44
$MASE_2$	0.43	0.17
MAE/Mean $ratio_1$	0.50	0.83
MAE/Mean $ratio_2$	0.81	0.32

assumed that the forecast values were determined using well-defined forecast methods.

For the two sets of forecast values, MAAPE was compared with MAPE, sMAPE, MASE, and the MAE/Mean ratio (see Section 1 for details of these measures). Table 2 summarizes the results of the five accuracy measures for the two forecasts, F_1 and F_2. As has been noted, MAPE cannot be defined unless data points with $A_t = 0$ are excluded, which indicates that MAPE is meaningless for

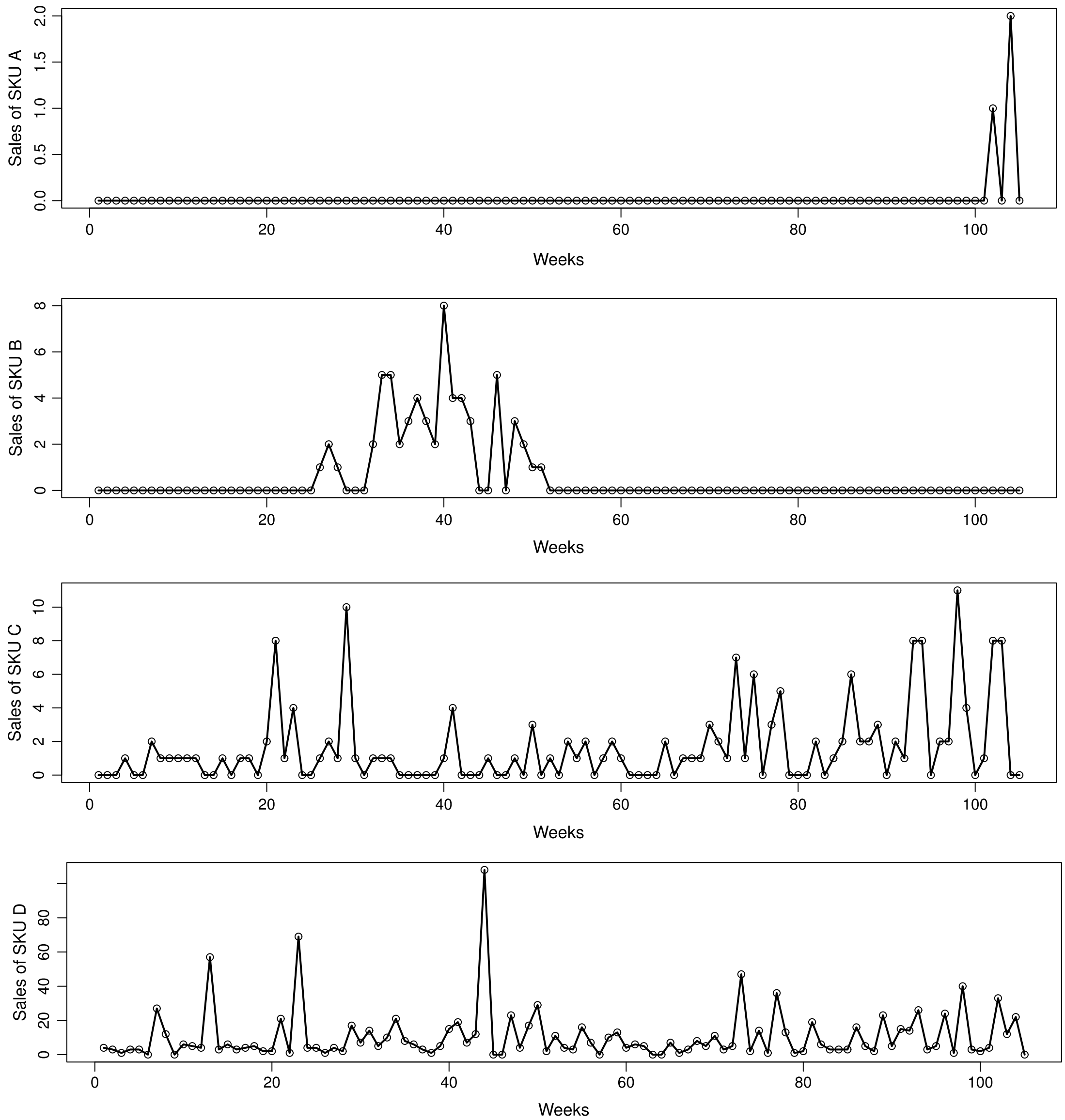


Fig. 9. Intermittent demand patterns of four SKUs in a specific store.

low volume data or data with periods of zero demand. If the points with $A_t = 0$ are excluded, MAPE_1 and MAPE_2 (which denote the MAPE values for F_1 and F_2, respectively) are 0.39 and 0.41 using the values of four months ($t = 28, 29, 32, 34$), although it is inevitable that the information from the other eight months be lost. In this case, MAPE determines that F_1 is a better forecast than F_2, which might not be an appropriate decision in this scenario. This is a result of MAPE placing a significantly heavier penalty on the positive outlier (i.e., the 28th month) in F_2 than necessary. In contrast, MAAPE obtains a finite value even when zero actual values are included: MAAPE_1 and MAAPE_2 are 1.17 and 1.15, respectively. As Table 2 shows, MAAPE makes a consistent decision, regardless of whether the data points for zero actual values are included or not: MAAPE determines that F_2 is a better forecast than F_1. This indicates that MAAPE is more robust to outliers than MAPE. sMAPE behaves similarly to MAAPE, consistently selecting F_2 as a better forecast than F_1. In contrast, MASE and the MAE/Mean ratio select F_2 if the points with $A_t = 0$ are included, while they select F_1 if the points with $A_t = 0$ are excluded.

5.2. Example 2: weekly retail sales data

A real dataset of retail sales was used to compare MAAPE with MAPE, sMAPE, MASE, and the MAE/Mean ratio. The data consisted of numerous time series for several stock keeping units (SKUs) belonging to a given category of products in a specific store of a large retail chain in the USA. Each time series consisted of weekly sales counts for the 105 weeks from October 29, 2005 to October 27, 2007. Fig. 9 illustrates the demand patterns of the four selected SKUs. From the top to the bottom of Fig. 9, the degrees of

Table 3
Forecast error measures for retail sales.

SKU	Measures	M1		M2		M3		M4	
		In	Out	In	Out	In	Out	In	Out
	MAPE	0.00	0.20	0.00	0.20	0.00	0.20	0.00	0.20
	MAAPE	0.00	0.16	0.00	0.16	0.00	0.16	0.00	0.16
A	sMAPE	0.00	0.40	0.00	0.40	0.00	0.40	0.00	0.40
	MASE	0.00	Undefined	0.00	Undefined	0.00	Undefined	0.00	Undefined
	MAE/Mean ratio	0.00	Undefined	0.00	Undefined	0.00	Undefined	0.00	Undefined
	MAPE	∞	∞	∞	∞	∞	∞	∞	∞
	MAAPE	1.33	1.57	0.93	1.57	1.33	1.57	0.93	1.57
B	sMAPE	1.72	2.00	1.20	2.00	1.72	2.00	1.23	2.00
	MASE	0.97	0.00	0.95	0.00	0.97	0.00	2.05	2.59
	MAE/Mean ratio	0.74	0.00	0.72	0.00	0.74	0.00	1.56	1.98
	MAPE	∞	∞	∞	∞	∞	∞	∞	∞
	MAAPE	0.86	0.84	0.86	0.85	0.93	0.90	0.89	0.97
C	sMAPE	1.13	1.11	1.16	1.11	1.21	1.06	1.17	1.10
	MASE	0.73	2.01	0.74	2.01	0.77	2.16	0.88	2.04
	MAE/Mean ratio	0.91	2.51	0.92	2.51	0.96	2.70	1.10	2.55
	MAPE	∞	∞	∞	∞	∞	∞	∞	∞
	MAAPE	0.84	0.89	0.75	0.90	0.84	0.89	0.78	0.89
D	sMAPE	0.91	1.09	0.85	1.09	0.91	1.09	0.92	1.12
	MASE	0.72	0.86	0.66	0.86	0.72	0.86	0.87	1.12
	MAE/Mean ratio	0.90	1.08	0.83	1.08	0.90	1.08	1.09	1.40

intermittence of the time series are 98%, 75%, 38%, and 7%, respectively, where the degree of intermittence is calculated as the number of time points with no sales divided by the total number of time points. Using these time series, the forecast accuracies of the following four forecast methods are compared: exponential smoothing state space models (Hyndman, Koehler, Snyder, & Grose, 2002) (denoted by M1), Holt–Winters (Holt, 2004) (denoted by M2), ARIMA (Hyndman & Khandakar, 2008) (denoted by M3), and Croston (Croston, 1972) (denoted by M4). These four forecast methods were implemented using the R package called 'forecast' (Hyndman, 2014) in R version 3.0.2. Like Hyndman (2006), we compare both the *in-sample* performances of these methods, by varying the origin and generating a sequence of one-period-ahead forecasts, and the *out-of-sample* performances, based on forecasting the data in the hold-out period using the information in the fitting period alone. For each time series, the data from the first 95 weeks are used as the in-sample data, while the remaining data from the 96th to 105th weeks are used as out-of-sample data.

For the four SKUs depicted in Fig. 9, Table 3 summarizes the forecast errors measured by the five metrics above for comparing the in-sample and out-of-sample performances of the four forecast methods. Note that we assume that if both the actual and forecast values are zero, each error measure is calculated to be zero regardless of the value of the denominator. In Table 3, the MAPE shows many infinite values, as a result of divisions by zero. The other four error measures, which were all aimed at overcoming this problem of MAPE, did not yield infinite values. However, if all of the actual values during the in-sample period were zeros, MASE and the MAE/Mean ratio yielded undefined values for the out-of-sample forecasts. MAAPE and sMAPE always provided reasonable results for all four forecasting methods.

5.3. *Example 3: simulated data*

Using a simulated example, MAPE and MAAPE were compared in terms of their abilities to select the appropriate forecast method. We generated actual demands, denoted by A_t, $t = 1, \ldots, N$, where N is the number of time points, from the negative binomial distribution in Eq. (4.4), with $r = 2$ and $p = 2/3$. For the same time period, we assumed that there were two forecasts available, denoted by F_{1t} and F_{2t}, $t = 1, \ldots, N$, which were obtained from two different forecast methods. Here, we assume that one forecast method is clearly superior to the other through the following setting:

$$F_{1t} = A_t + \epsilon_{1t}, \quad t = 1, \ldots, N,$$
$$F_{2t} = A_t + \epsilon_{2t}, \quad t = 1, \ldots, N,$$

where $\epsilon_{1t} \sim N(0, 0.1^2)$ and $\epsilon_{2t} \sim N(0, 0.2^2)$; thus, with the small variance term, F_1 is designed to be a better forecast than F_2. First, we generated a sequence of actual demands; for the generated actual data, 100 pairs of the two sequences of forecast values were generated as above, with $N = 10$. Using both the forecast data of each pair and the actual data, MAPE and MAAPE values were calculated for the two forecast methods F_1 and F_2, based on the forecast method that was selected by each measure. Then, the performances of MAPE and MAAPE were compared based on their error rates: the number of incorrect selections divided by the total number of trials (=100). That is, the error rate indicated how often F_2 was selected as being better than F_1, even though F_1 was designed to be better than F_2. When calculating the APE with $A = 0$, a small value of 0.01 was added to the actual value A in order to avoid the problem of division by zero. This procedure was then repeated 1000 times, and the distributions of the 1000 error rates for MAPE and MAAPE thus obtained were compared in Fig. 10, using boxplots. In terms of error rates, MAAPE outperformed MAPE; that is, it selected F_1 as the better forecast more often.

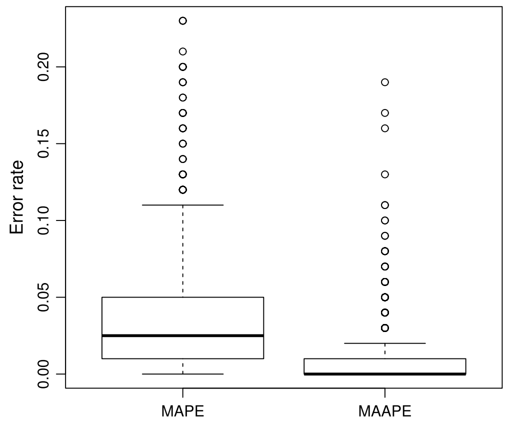


Fig. 10. Boxplots of 1000 error rates for MAPE and MAAPE.

6. Conclusion

We have developed a novel accuracy measure called the mean arctangent absolute percentage error (MAAPE) by modifying MAPE, which is the most popular accuracy measure. As the proposed measure is a MAPE that has been transformed using the arctangent (inverse tangent) function, it inherently preserves the advantages of MAPE; thus, MAAPE is scale-independent, can be interpreted intuitively as an absolute percentage error, and is simple to calculate. In addition, the bounded range of the arctangent function allows the MAAPE to overcome the MAPE's limitation of going to infinity as the actual value goes to zero. We visualized MAAPE in order to demonstrate its advantages over MAPE, and proved that AAPE is a concave-down function for $F > 2A > 0$, which enables MAAPE to obtain a more balanced penalty between positive and negative errors than MAPE, although the penalty function of MAAPE remains asymmetric. MAAPE is also more robust than MAPE due to the bounded influences of outliers; thus, MAAPE can be particularly useful when extremely large errors are due to mistaken or incorrect observations. However, if extremely large errors are considered as having important business implications, rather than merely being due to mistaken observations, MAAPE is not recommended. Via a simulation, we have also compared APE and AAPE based on their optimal point predictions under three strategies, and have demonstrated that MAAPE is less biased than MAPE. Our applications to real and simulated data have also demonstrated the effectiveness of MAAPE in practice.

Acknowledgment

This research was partially supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (2015R1C1A1A02037090).

Appendix A. Proof of Theorem 1

$$\frac{\partial^2}{\partial^2 A} AAPE = \frac{\partial^2}{\partial^2 A} \arctan\left(\frac{F-A}{A}\right) \quad \text{for } 0 < A < F$$

$$= \frac{\partial}{\partial A}\left(\frac{-F}{A^2+(F-A)^2}\right)$$

$$= \frac{2F}{(A^2+(F-A)^2)^2}(2A-F).$$

Appendix B. Proof of Proposition 1

Suppose that A_L is realized. Then, APE can be reduced by decreasing F. However, under the outcome A_H, decreasing F implicitly means increasing APE. Thus, equilibrium occurs when the APE under outcome A_L and the APE under outcome A_H have been equalized. That is,

$$APE(A_L, F_{APE}^{(3)}) = APE(A_H, F_{APE}^{(3)}),$$

which is

$$\left|\frac{F_{APE}^{(3)} - A_L}{A_L}\right| = \left|\frac{A_H - F_{APE}^{(3)}}{A_H}\right|. \tag{B.1}$$

For the same reason, equilibrium occurs when the AAPE under outcome A_L and the AAPE under outcome A_H are equalized:

$$\arctan\left|\frac{F_{AAPE}^{(3)} - A_L}{A_L}\right| = \arctan\left|\frac{A_H - F_{AAPE}^{(3)}}{A_H}\right|. \tag{B.2}$$

Because arctan is a one-to-one function, Eq. (B.2) is equivalent to Eq. (B.1). Thus,

$$F_{APE}^{(3)} = F_{AAPE}^{(3)}.$$

References

Armstrong, J., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons with discussion. *International Journal of Forecasting*, *8*, 69–80.

Bowerman, B. L., O'Connell, R. T., & Koehler, A. B. (2004). *Forecasting, time series and regression: An applied approach*. Belmont, CA: Thomson Brooks/Cole.

Byrne, R. F. (2012). Beyond traditional time-series: Using demand sensing to improve forecasts in volatile times. *The Journal of Business Forecasting*, *31*, 13.

Chatfield, C. (1988). Apples, oranges and mean square error. *International Journal of Forecasting*, *4*, 515–518.

Croston, J. D. (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly*, 289–303.

Fildes, R., & Makridakis, S. (1988). Loss functions and forecasting. *International Journal of Forecasting*, *4*, 545–550.

Hanke, J. E., & Reitsch, A. G. (1995). *Business forecasting* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*, 5–10.

Hyndman, R. J. (2006). Another look at forecast accuracy metrics for intermittent demand. *Foresight: The International Journal of Applied Forecasting*, *4*, 43–46.

Hyndman, R.J. (2014). Forecast: Forecasting functions for time series and linear models, R package version, 5.4.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *26*.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*, 439–454.

Kolassa, S., & Schütz, W. (2007). Advantages of the MAD/Mean Ratio over the MAPE. *Foresight: The International Journal of Applied Forecasting*, *6*, 40–43.

Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, *9*, 527–529.

Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., & Lewandowski, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111–153.

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society. Series A*, *142*, 97–145.

Makridakis, S., Wheelwright, S., & Hyndman, R. J. (1998). *Forecasting: Methods and applications* (3rd ed.). New York: John Wiley and Sons.

McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, *113*, 259–262.

Morlidge, S., et al. (2015). Measuring the quality of intermittent-demand forecasts: It's worse than We've thought!. *Foresight: The International Journal of Applied Forecasting*, 37–42.

Swanson, D. A., Tayman, J., & Barr, C. F. (2000). A note on the measurement of accuracy for subnational demographic estimates. *Demography*, *2*, 193–201.

Syntetos, A. (2001). *Forecasting of intermittent demand*. (Ph.D. thesis), Buckinghamshire Chilterns University College, Brunel University, UK.

Syntetos, A., Lengu, D., & Babai, M. Z. (2013). A note on the demand distributions of spare parts. *International Journal of Production Research*, *51*, 6356–6358.

Teunter, R. H., & Duncan, L. (2009). Forecasting intermittent demand: a comparative study. *Journal of the Operational Research Society*, *60*, 321–329.

Theil, H. (1966). *Applied economic forecasting*. Chicago, IL: Rand McNally.

Thompson, P. A. (1990). An MSE statistic for comparing forecast accuracy across series. *International Journal of Forecasting*, *6*, 219–227.

Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, *66*, 1352–1362.

Sungil Kim, who is currently working as a data scientist for Samsung SDS, received his Ph.D. degree in Industrial Engineering, M.S. degrees in Industrial Engineering and Statistics from the Georgia Institute of Technology, and B.S. degree in Industrial Engineering from Yonsei University. His major research interests are in the areas of demand forecasting, data mining and machine learning, multi-level multi-scale modeling, and business analytics.

Heeyoung Kim is currently an assistant professor in Department of Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology (KAIST). She received her Ph.D. degree in Industrial Engineering from the Georgia Institute of Technology, and B.S. and M.S. degrees in Industrial Engineering from Korea Advanced Institute of Science and Technology (KAIST). Her research interests are in the areas of data mining and machine learning, nonparametric statistical methods, and business applications with spatio-temporal data analysis.