Using the R-MAPE index as a resistant measure of forecast accuracy

Article *in* Psicothema · November 2013

DOI: 10.7334/psicothema2013.23 · Source: PubMed

4 authors:

Juan José Montaño
University of the Balearic Islands
50 PUBLICATIONS **1,062** CITATIONS

Alfonso Palmer
University of the Balearic Islands
106 PUBLICATIONS **1,537** CITATIONS

Albert Sesé
University of the Balearic Islands
84 PUBLICATIONS **1,236** CITATIONS

Berta Cajal
University of the Balearic Islands
27 PUBLICATIONS **235** CITATIONS

Psicothema

Psicothema 2013, Vol. 25, No. 4, 500-506
doi: 10.7334/psicothema2013.23

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

Using the R-MAPE index as a resistant measure of forecast accuracy

Juan José Montaño Moreno, Alfonso Palmer Pol, Albert Sesé Abad and Berta Cajal Blasco
Universidad de las Islas Baleares

Abstract

Background: The mean absolute percentage error (MAPE) is probably the most widely used goodness-of-fit measure. However, it does not meet the validity criterion due to the fact that the distribution of the absolute percentage errors is usually skewed to the right, with the presence of outlier values. In these cases, MAPE overstates the corresponding population parameter. In this study, we propose an alternative index, called Resistant MAPE or R-MAPE based on the calculation of the Huber M-estimator, which allows overcoming the aforementioned limitation. **Method:** The results derived from the application of Artificial Neural Network (ANN) and Autoregressive Integrated Moving Average (ARIMA) models are used to forecast a time series. **Results:** The arithmetic mean, MAPE, overstates the corresponding population parameter, unlike R-MAPE, on a set of error distributions with a statistically significant right skew, as well as outlier values. **Conclusions:** Our results suggest that R-MAPE represents a suitable alternative measure of forecast accuracy, due to the fact that it provides a valid assessment of forecast accuracy compared to MAPE.

Keywords: Time series, error measures, outliers, neural networks, ARIMA models.

Resumen

El índice R-MAPE como medida resistente del ajuste en la previsión. **Antecedentes:** el Promedio del Error Porcentual Absoluto (MAPE) es probablemente la medida de adecuación de la previsión más ampliamente utilizada. Sin embargo, no cumple el criterio de validez debido a que la distribución de los errores porcentuales absolutos habitualmente presenta una forma asimétrica a la derecha con presencia de valores alejados. En estos casos, el MAPE proporciona una sobreestimación del correspondiente parámetro poblacional. En el presente trabajo se propone un índice alternativo, denominado MAPE Resistente o R-MAPE, y basado en el cálculo del M-estimador de Huber, el cual permite superar la mencionada limitación. **Método:** se utilizan los resultados derivados de la aplicación de modelos de Red Neuronal Artificial (ANN) y modelos Autorregresivos Integrados de Media Móvil (ARIMA) en la previsión de una serie temporal. **Resultados:** se puede observar que la media aritmética, el MAPE, realiza una sobreestimación del correspondiente parámetro poblacional, a diferencia del R-MAPE, sobre un conjunto de distribuciones de errores con asimetría a la derecha y presencia de valores alejados. **Conclusiones:** nuestros resultados ponen de manifiesto que el R-MAPE representa una adecuada alternativa en la medición del ajuste en la previsión, debido a que proporciona una evaluación válida de dicho ajuste en comparación al MAPE.

Palabras clave: series temporales, medidas de error, valores alejados, redes neuronales, modelos ARIMA.

In recent decades, numerous comparative studies have been carried out with the aim of identifying the most accurate method for time series forecasting (Armstrong & Collopy, 1992; Palmer, Montaño, & Franconetti, 2008). This is mainly due to the fact that obtaining accurate forecasts has become a crucial issue for researchers, practitioners and policy makers in a range of disciplines such as tourism, economics or industry.

Despite the consensus on the need to develop accurate forecasts and the recognition of their corresponding benefits, there is no one model that stands out in terms of forecasting accuracy (Law & Au, 1999). In this respect, one of the most widely used procedures in time series forecasting is the Box-Jenkins methodology (Box & Jenkins, 1976), which is based on the fit of a special type of linear statistical model known as ARIMA (Autoregressive Integrated Moving Average).

In recent years, several alternative methods to the traditional ones have been proposed in order to carry out time series forecasting. This is the case of Artificial Neural Networks (ANN) which has aroused great interest in fields as diverse as biology, psychology, medicine, economics, mathematics, statistics and computers. The reason behind this interest is that ANN are universal function approximators capable of mapping any linear or non-linear function (Cybenko, 1989; Funahashi, 1989; Hornik, Stinchcombe, & White, 1989; Wasserman, 1989). Due to their flexibility in function approximation, ANN are powerful methods in tasks involving pattern classification, estimating continuous variables and forecasting (Kaastra & Boyd, 1996).

Nevertheless, studies comparing the time series forecasting abilities of traditional methods and neural networks (e.g., Balestrino, Bini Verona, & Santanche, 1994; Foster, Collopy, & Ungar, 1992; Pattie & Snyder, 1996; Palmer, Montaño, & Franconetti,

Received: January 21, 2013 • Accepted: May 28, 2013
Corresponding author: Juan José Montaño Moreno
Facultad de Psicología
Universidad de las Islas Baleares
07122 Palma de Mallorca (Spain)
e-mail: juanjo.montano@uib.es

2008; Sharda & Patil, 1990; Tang, Almeida, & Fishwick, 1991) have mixed results. Some favour traditional forecasting methods while others favour neural networks. One of the factors that may explain this fact in part is, as Makridakis, Wheelwright and McGee (1983) indicate, because there is no single universally accepted measure of accuracy. The selection of appropriate error measures in forecasting is always a problem because, as Mathews and Diamantopoulos (1994) point out, no single measure gives an unambiguous indication of forecasting performance, while the use of multiple measures makes comparisons between forecasting methods difficult and unwieldy. Finally, research findings indicate that the performance of different methods depends upon the accuracy measure used (Makridakis, 1993).

The mean absolute percentage error (MAPE) is probably the most widely used forecasting accuracy measurement (Armstrong & Collopy, 1992; Goodwin & Lawton, 1999; Ren & Glasure, 2009). MAPE has important, desirable features including reliability, unit-free measure, ease of interpretation, clarity of presentation, support of statistical evaluation, and the use of all the information concerning the error. However, several authors (Armstrong & Collopy, 1992; Makridakis, 1993) have questioned its validity, due to the characteristics of error distributions.

This study aims to analyse the statistical properties of the most widely used measure in forecast error estimation, the mean absolute percentage error (MAPE), and to propose an alternative index, called Resistant MAPE or R-MAPE, which makes it possible to overcome the limitations detected in this measure. Thereby, the results derived from the application of ANN and ARIMA models in time series forecasting of electrical energy consumption were used.

Analysis of MAPE index properties

According to the National Research Council (1980), any summary measure of error must meet five basic criteria: measurement validity, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation. In attempting to meet these criteria, the summary measure of population forecast error most often used is MAPE, the mean absolute percentage error (Ahlburg, 1995; Isserman, 1977; Murdock, Leistritz, Hamm, Hwang, & Parpia, 1984; Smith, 1987; Smith & Sincich, 1990, 1992; Tayman, Schafer, & Carter, 1998). MAPE obeys the following mathematical expression:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{y_t}\times 100$$

where *n* is the size of the sample, $\hat{y}_t$ is the value predicted by the model for time point *t*, and y_t is the value observed at time point *t*.

Meanwhile, Lewis (1982) drew up a table (see Table 1) containing typical MAPE values for industrial and business data and their interpretation.

Table 1
Interpretation of typical MAPE values

MAPE	Interpretation
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

Source: Lewis (1982, p. 40)

According to several authors (Tayman & Swanson, 1999), MAPE satisfactorily meets at least four of the aforementioned criteria, but is less satisfactory in meeting the validity criterion when used in evaluating the accuracy of population forecasts. Noncompliance with the validity criterion is due to the fact that the absolute percentage error distribution —characterised by having only positive values with no upper bound—usually has a right or positive skew brought about by the presence of outlier values to this side of the distribution. In other words, most observations in a right-skewed distribution fall toward the lower values (Fildes, 1992), and a relatively few large values or outliers form a tail that slopes to the right. As a result, in these cases the arithmetic mean of the percentage error calculated in a sample provides an overstatement of the corresponding population mean. If MAPE overstates forecast error, it is not valid, in a criterion-related sense (Carmines & Zeller, 1979), for evaluating the accuracy of population forecasts. Criterion-related validity is defined as the degree of correspondence between a given measure and some phenomenon of interest that is external to the measure itself; the latter is the criterion (Swanson, Tayman, & Barr, 2000).

In view of the aforementioned limitation of MAPE, in recent years several solutions have been proposed. Hence, Makridakis (1993) suggests the elimination of errors declared as outliers in calculating the index. Later on, Makridakis and Hibon (1995) use a sort of linear transformation of MAPE called symmetrical MAPE or SMAPE. Nevertheless, the empirical results obtained by Tayman and Swanson (1999) indicate that SMAPE does not constitute an adequate alternative, due to the fact that the data transformed are difficult to interpret and the index obtained is still affected by the presence of outlier values just like MAPE. A similar problem is presented by MAPE-T (MAPE-Transformed) (Swanson, Tayman, & Barr, 2000), calculated through the use of a modified Box-Cox method (Box & Cox, 1964). With the aim of overcoming the limitations of MAPE-T, Swanson, Tayman and Barr (2000) propose MAPE-R (MAPE-Rescaled), a procedure to convert MAPE-T into the same scale as the original observations. However, the calculation process until reaching MAPE-R is complex and requires the user to possess skills in the field of statistical modelling.

In this study we propose, as the most satisfactory and simplest solution to obtain, the calculation of M-estimators in order to obtaining a location index of the absolute percentage error distribution which is resistant or insensitive to the presence of outlier values.

The R(esistant)-MAPE index

The arithmetic mean provides a value of the variable which represents the centre of gravity of the distribution (in which the distribution of observations is balanced). Nevertheless, it is worth remembering that the use of this classical location index should be limited only to those occasions on which the distribution of the variable is symmetrical.

As opposed to the arithmetic mean, there are other more appropriate measures of central tendency to describe data when dealing with asymmetrical distributions and/or with outlier values. In this sense, we have indexes such as the trimmed mean which consists of eliminating a proportion of the data from each extreme and calculating the mean of the remaining values, or the Winsored mean (Miller, 1986), which instead of eliminating a whole number of cases from each extreme, substitutes them for the last value, at each extreme, which is part of the analysis. Precisely, Armstrong and Collopy (1992) propose reducing the effect of outlier errors by trimming or Winsorizing in time series forecasting.

On the other hand, Exploratory Data Analysis, generally known as EDA (Tukey, 1977) offers a set of simple, resistant and clear techniques. EDA, contrary to traditional descriptive analysis, places more relevance on resistant measures and on graphic information. As a result, EDA incorporates indexes and graphics that overcome the problems presented by classical descriptive statistics when facing non symmetrical distributions and the presence of outlier values.

Among the indexes included in EDA we find the median and the M-estimators (Huber, 1964). The median is defined as the value of the variable that divides the distribution into two equal parts, each of which contains 50% of the observations, whereas the M-estimators look for a location index from the total set of observations, by pondering these depending on how near or far they are from the centre of the data.

An estimator of location or of scale is said to be resistant if slight changes in the distribution of the data have hardly any effect on its value. From this point of view, it is obvious that the introduction of only one extreme value in the distribution means there is a change in the arithmetic mean of that distribution. Thus, the arithmetic mean and, therefore, MAPE, is not a resistant or valid index.

However, in order to be able to talk about resistance there is a series of properties from which it is possible to establish what the best estimator for representing the absolute percentage error distribution is. In this sense, we will focus on the search for the best location index, basically in the presence of extreme values, that is, the so-called 'outliers'.

Below, we briefly define the properties to be taken into account (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Huber, 1981; Wilcox, 1997):

1. The *influence function* determines the influence an anomalous value has on the value of the estimator. If the influence function is not bounded, it means that the further away the anomalous datum is, the greater the influence exerted on the estimator. This is what happens in the arithmetic mean of the absolute percentage error (MAPE), whose influence function is linear and is not therefore bounded.
2. The *gross-error sensitivity* measures the influence exerted by a certain quantity of contamination (anomalous values) in the data on the value of the estimator. If this value is finite, the estimator is said to be B-robust.
3. The *local-shift sensitivity* is the one determined by small fluctuations in the data, and it is desirable for it to be small and finite.
4. Under the strategy that it is convenient to eliminate clearly anomalous values, the influence function must be zero from a certain value. For symmetrical distributions around zero, the *rejection point* is the value from which the data must be rejected. It is desirable for the estimator to have a finite rejection point.
5. The *breakdown point* of an estimator is the percentage of outliers the estimator can stand before breaking down, that is, before ceasing to be valid, and this defines the quantitative robustness. An estimator is resistant only if its breakdown point is greater than zero.

The M-estimators of location weight the observations on the basis of their relative distance from the centre of the distribution, whereas a Winsorized mean replaces a predetermined a percentage of observations with a non-outlier value, and what a trimmed mean does is to eliminate this percentage of observations. Why use an M-estimator and not a Winsorized mean or a trimmed mean, instead of an arithmetic mean?

The sample median is B-robust, its gross-error sensitivity is finite, its local-shift sensitivity is infinite, its rejection point is infinite, it is qualitatively robust and its breakdown point is ½.

The Winsorized mean is B-robust; however it has an infinite value for local-shift sensitivity, when what is desirable is for the estimator to have the smallest possible finite value. Likewise, it has a non-finite rejection point and an a value breakdown point, which means that it stands a maximum proportion of alpha value of outliers in order to continue making sense as a location index. Lastly, the Winsorized mean is not a qualitatively robust estimator.

The trimmed mean is B-robust, therefore its gross-error sensitivity is bounded and its local-shift sensitivity is finite, depending on the alpha value, in which case, in this sense, it is an improvement on the Winsorized mean. However, the rejection point is still infinite and the breakdown point is alpha, just like the Winsorized mean.

Huber's M-estimator is an estimator with good properties, both in terms of resistance and accuracy, since, amongst others, it is qualitatively robust and reaches the maximum possible breakdown point and is the most optimum B-robust estimator, that is to say, its gross-error sensitivity is bounded. Despite the fact that Huber's estimator does not have a finite rejection point, as Hampel's estimator (three part redescending) may have, this does not have the efficiency properties that Huber's estimator has.

Hence, of the different M-estimators, Huber's is one of the ones that provides values nearest the arithmetic mean, due to the fact that it weights with a value of 1 a greater quantity of central data than other M-estimators, Tukey or Andrews types, which weight all observations below 1, in which case their comparison with the arithmetic mean is one of the most conservative.

One advantage of the M-estimator over a trimmed mean is that this will always eliminate data from both extremes of the distribution, also eliminating possible valid data, whereas the M-estimator will focus more on the extreme with most outlier values, and could even leave the other extreme of the distribution without modification.

The sample mean, m, as an estimator of central tendency, is not B-robust, its gross-error sensitivity is infinite, its local-shift sensitivity is 1, its rejection point is infinite, it is not qualitatively robust and its breakdown point is zero. Automatically choosing the arithmetic mean as the best index of location, and efficiency only makes sense when the population distribution is the normal distribution, but this does not show any protection against the presence of outliers.

Despite the drawbacks of the arithmetic mean in the absolute percentage error distribution, and in spite of the existence of simple procedures with the advantages already outlined, this is still used practically exclusively as if it were the only available location estimator. In this study we propose the Resistant MAPE or R-MAPE index based on the use of Huber's M-estimator as an appropriate alternative as opposed to the arithmetic mean, in order to represent the absolute percentage error distribution in forecasting time series.

Method

Data

In this study we used the data concerning the total monthly electrical energy consumption (MWh unit) in the Balearic Islands between January 1983 and April 2003, obtaining a time series made up of 244 time points (from x_1 to x_{244}). In this sense, the forecast of electrical consumption constitutes one of the most paradigmatic problems in the field of time series analysis (Pao, 2006).

Figure 1A shows the graphical representation of the original time series. Following the traditional procedures of pre-processing

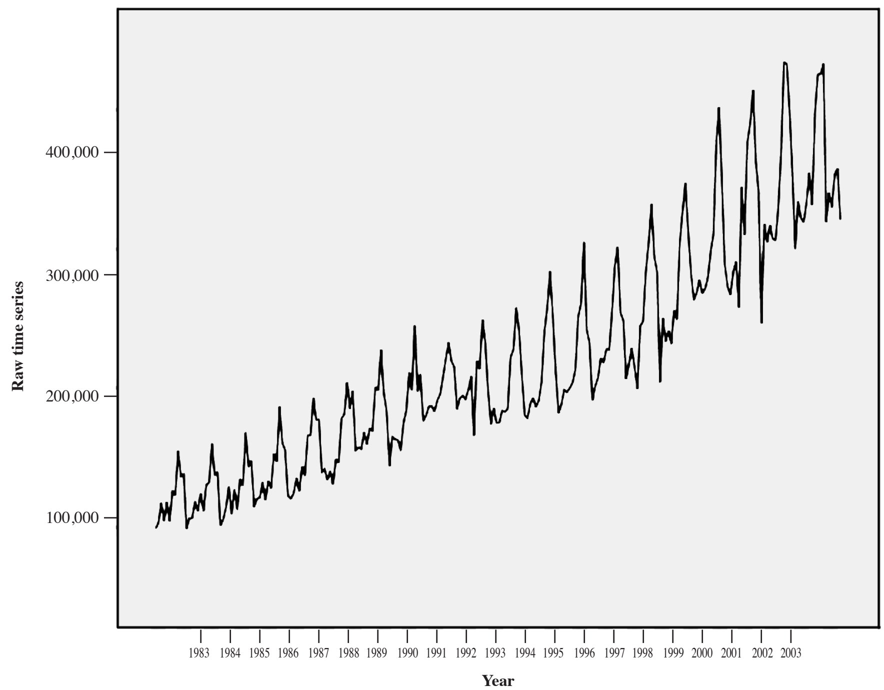

Figure 1A. Raw time series

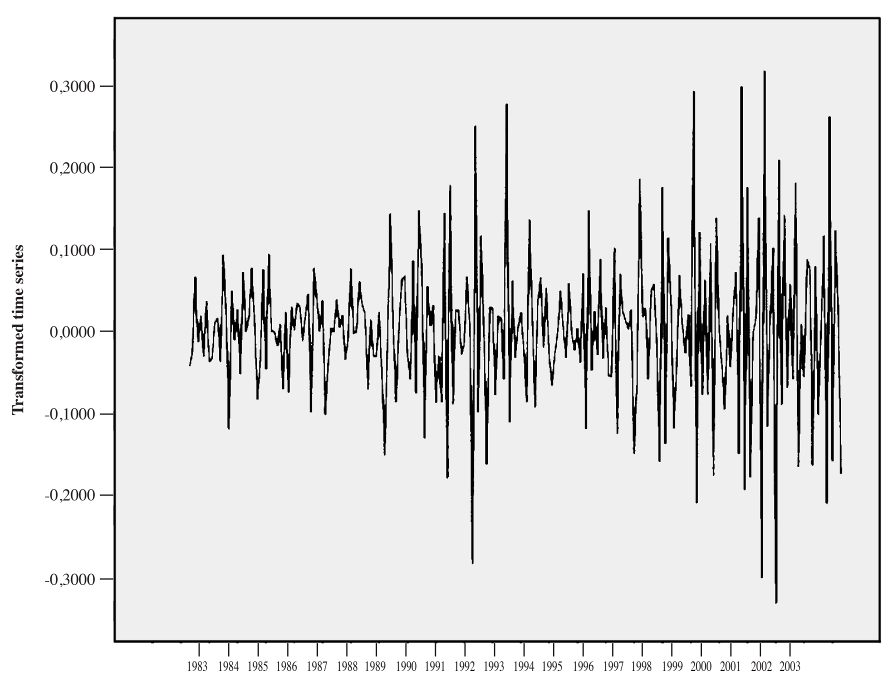

Figure 1B. Transformed time series

Figure 1. *Graphic representation of the original and transformed time series*

time series, a logarithmic transformation and two differentiations —one of order 1 and the other of order 12—were applied. Figure 1B shows the time series after applying these transformations.

Then, for the neural network model design, the set of patterns was divided into three groups: the training group, made up of the time points corresponding to the period between 1983 and 1996; the validation group, made up of the time points corresponding to the period between 1997 and 1999; and, finally, the test group, made up of the time points corresponding to the period between 2000 and 2003.

Meanwhile, for the ARIMA model design, the period between 1983 and 1999 was used to estimate the parameters, and the period between 2000 and 2003 was used for the accuracy test in order to compare with the neural network models.

Forecasting models

Artificial Neural Networks (ANNs) are information processing systems whose structure and workings are inspired by biological neural networks. They consist of a large number of simple processing elements called nodes or neurons which are arranged in layers. Each neuron is connected to other neurons through communication links, each of which has an associated weight. The knowledge the ANN has concerning a certain task is found in the weights.

ANN can be considered general, flexible, nonlinear statistical techniques capable of learning complex relationships between variables in a multitude of fields of study. In this study, we used four network models which have proven to be appropriate in time series forecasting (for a more detailed description of these models, consult Montaño, Palmer, & Muñoz, 2011): Multilayer Perceptron (MLP), Radial Basis Function (RBF), Generalized Regression Neural Network (GRNN) and Recurrent Neural Networks (RNN). With the aim of comparing the accuracy of network models with a classical model, we also applied the ARIMA model as it is the most widely used procedure for time series forecasting.

Results

Table 2 shows the descriptive analysis of the absolute percentage error (APE) distribution in the test set (n = 40) for each of the five forecasting models analysed. More specifically, we provide R-MAPE based on Huber's M-estimator, the median, MAPE, the ratio between MAPE and the median, the robust variation coefficient, 95% confidence interval of the asymmetry parameter, the Shapiro-Wilk normality test value and, finally, the maximum error observed. Meanwhile, Figure 2 represents the corresponding Box-Plot diagrams. The results show that, in all cases, the error distributions do not follow the normal curve; they have a statistically significant right skew, as well as outlier values.

The error distribution shape identified and the presence of outlier values means that the arithmetic mean, MAPE, is not an appropriate index, overstating the corresponding population parameter. In this sense, it can be observed that the ratio between MAPE and the median oscillates between 7% (GRNN) and 38% (ARIMA), showing that MAPE has a systematic tendency to overstate forecast error. Although the correspondence is not perfect, the overstatement of the error by MAPE tends to increase as the degree of asymmetry increases. For instance, the model which has the greatest skewness, the ARIMA model with a skewed confidence interval between 1.07 and 2.92, is the one which has the greatest ratio between MAPE and the median, with 38%. Meanwhile, R-MAPE provides lower values which are nearer the value of the median, in comparison to the value of the arithmetic mean. This is due to the fact that Huber's M-estimator is not influenced by outlier values, focusing its attention on the central body of the error distribution.

With respect to forecasting accuracy, all the models analysed show a good fit to the test data, with the ARIMA model clearly superior in comparison to the neural network models. The good forecasting results improve if, instead of using MAPE, we use R-MAPE. Thus, for instance, if we take the arithmetic mean (MAPE) as a basis, the RBF and GRNN models would be considered simply as good forecasting models, following the categorization of Lewis (1982). On the other hand, if we take Huber's M-estimator value (R-MAPE) as the basis, all the models analysed would be considered highly accurate forecasting models.

Discussion

In our study we propose the use of the R-MAPE index, as an alternative to MAPE, based on the calculation of Huber's M-estimator. We were able to analyse, from a theoretical point of view, how this alternative complies satisfactorily with a series of statistical properties compared to other estimators, in terms of validity, resistance and accuracy. It is important to point out that R-MAPE maintains the properties of MAPE, that is, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation, and overcomes its limitation in terms of validity criteria.

The empirical results obtained in the study reveal that, in all cases, the error distribution is right skewed with the presence of outlier values, leading to MAPE overstating the corresponding

Table 2
Descriptive analysis of the absolute percentage error distribution

Model	R-MAPE	Median APE	MAPE	Ratio MAPE to median	Robust coefficient of variation	CI 95% skewness	Shapiro-Wilk Test	Maximum value
MLP	7.35	7.22	8.45	1.17	0.42	0.34; 1.88	0.91**	27.12
RBF	8.93	8.24	10.01	1.21	0.51	0.16; 1.70	0.93 *	32.04
RNN	7.10	6.11	8.10	1.32	0.50	0.24; 1.78	0.91**	24.46
GRNN	9.53	9.49	10.17	1.07	0.59	0.23; 1.77	0.92 *	35.20
ARIMA	3.64	3.42	4.74	1.38	0.44	1.07; 2.92	0.78**	19.15

Note: * <.05; ** <.01

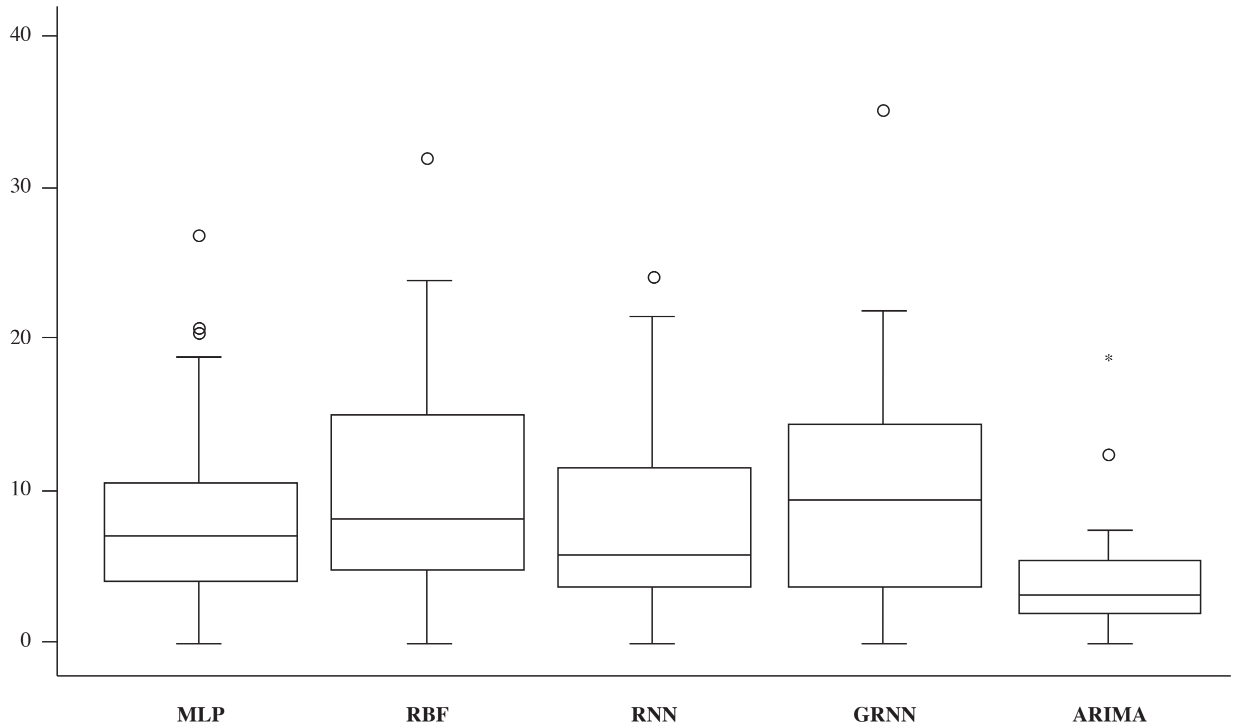

Figure 2. *Box-Plot representation of the absolute percentage error distribution*

population parameter. This overstatement was observed in reference to the value of the median. Along these lines, our results suggest the Huber M-estimator represents a suitable alternative measure of forecast accuracy, due to the fact that it provides a valid assessment of forecast accuracy. Unlike MAPE, R-MAPE incorporates outlier information, but does not allow outliers to dominate the summary measure of error.

The comparative study conducted between neural network models and ARIMA models on the basis of the R-MAPE index revealed that all the models show highly accurate forecasting, with the ARIMA model as the best one. Therefore, it was possible to substantiate once again the fact that neural network models constitute a technique to be taken into account by researchers, practitioners and policy makers in forecasting time series.

Finally, among its contributions, this study aims to suggest that in all the fields of Psychology where time series models—classical or more innovators like ANN—have been applied, it would be preferable to use a valid error measures such as R-MAPE. By way of illustration, the fields of application could be about the use and abuse of psychoactive substances (Sears, Davis, Guydish, & Gleghorn, 2009), psychophysiological activity (Janjarasjitt, Scher, & Loparo, 2008), criminal or violent behaviour (Pridemore & Chamlin, 2006), assessment of psychological intervention programmes (Tschacher & Ramseyer, 2009), teaching methodologies (Escudero & Vallejo, 2000) or psychopathology (Valiyeva, Herrmann, Rochon, Gill, & Anderson, 2008).

References

Ahlburg, D. (1995). Simple versus complex models: Evaluation, accuracy, and combining. *Mathematical Population Studies, 5*, 281-290.

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

Balestrino, A., Bini Verona, F., & Santanche, M. (1994). Time series analysis by neural networks: Environmental temperature forecasting. *Automazione e Strummentazione, 42*, 81-87.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, 26*, 211-252.

Box, G., & Jenkins, G.M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden Day.

Carmines, E., & Zeller, R. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematical Control, Signal and Systems, 2*, 303-314.

Escudero, J.R., & Vallejo, G. (2000). Aplicación del diseño de series temporales múltiples a un caso de intervención en dos clases de Enseñanza General Básica. *Psicothema, 12*, 203-210.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting, 8*, 81-98.

Foster, W.R., Collopy, F., & Ungar, L.H. (1992). Neural network forecasting of short, noisy time series. *Computers and Chemical Engineering, 16*, 293-297.

Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural Networks, 2*, 183-192.

Goodwin, P., & Lawton R. (1999). On the asymetry of the symmetric MAPE. *International Journal of Forecasting, 15*, 405-408.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., & Stahel, W.A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley-Interscience.
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*, 359-366.
Huber, P. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 36*, 73-101.
Huber, P. (1981). *Robust statistics*. New York: Wiley.
Isserman, A. (1977). The accuracy of population projections for subcounty areas. *Journal of the American Institute for Planners, 43*, 247-259.
Janjarasjitt, S., Scher, M.S., & Loparo, K.A. (2008). Nonlinear dynamical analysis of the neonatal EEG time series: the relationship between sleep state and complexity. *Clinical Neurophysiology, 119*, 1812-1823.
Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing, 10*, 215-236.
Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management, 20*, 89-97.
Lewis, C.D. (1982). *Industrial and business forecasting methods*. London: Butterworths.
Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting, 9*, 527-529.
Makridakis, S., & Hibon, M. (1995). *Evaluating accuracy (or error) measures*. INSEAD Working Papers Series 95/18/TM. Fontainebleau, France.
Makridakis, S., Wheelwright, S.C., & McGee, V.E. (1983). *Forecasting: Methods and Applications*. New York: Wiley.
Mathews, B.P., & Diamantopoulos, A. (1994). Towards a taxonomy of forecast error measures: A factor-comparative investigation of forecast error dimensions. *Journal of Forecasting, 13*, 409-416.
Miller, R.G. (1986). *Beyond ANOVA. Basics of applied statistics*. Nueva Cork: John Wiley and Sons.
Montaño, J.J., Palmer, A., & Muñoz, P. (2011). Artificial neural networks applied to forecasting time series. *Psicothema, 23*, 322-329.
Murdock, S., Leistritz, L., Hamm, R., Hwang, S., & Parpia, B. (1984). An assessment of the accuracy of a regional economic-demographic projection model. *Demography, 21*, 383-404.
National Research Council (1980). *Estimating population and income for small places*. Washington, DC: National Academy Press.
Palmer, A., Montaño, J.J., & Franconetti, F.J. (2008). Sensitivity analysis applied to artificial neural networks for forecasting time series. *Methodology, 4*, 80-86.
Pao, H.T. (2006). Modeling and forecasting the energy consumption in Taiwan using artificial neural networks. *The Journal of American Academy of Business, 8*, 113-119.
Pattie, D.C., & Snyder, J. (1996). Using a neural network to forecast visitor behavior. *Annals of Tourism Research, 23*, 151-164.
Pridemore, W.A., & Chamlin, M.B. (2006). A time-series analysis of the impact of heavy drinking on homicide and suicide mortality in Russia, 1956-2002. *Addiction, 101*, 1719-1729.
Ren, L., & Glasure, Y. (2009). Applicability of the revised mean absolute percentage errors (MAPE) approach to some popular normal and non-normal independent time series. *International Advances in Economic Research, 15*, 409-420.
Sears, C., Davis, T., Guydish, J., & Gleghorn, A. (2009). Investigating the effects of San Francisco's treatment on demand initiative on a publicly-funded substance abuse treatment system: A time series analysis. *Journal of Psychoactive Drugs, 41*, 297-304.
Sharda, R., & Patil, R. (1990). Neural networks as forecasting experts: An empirical test. *Proceedings of the 1990 International Joint Conference on Neural Networks Meeting, 2*, 491-494.
Smith, S. (1987). Tests of accuracy and bias for county, population projections. *Journal of the American Statistical Association, 82*, 991-1003.
Smith, S., & Sincich, T. (1990). On the relationship between length of base period and population forecast errors. *Journal of the American Statistical Association, 85*, 367-375.
Smith, S., & Sincich, T. (1992). Evaluating the forecast accuracy and bias of alternative projections for states. *International Journal of Forecasting, 8*, 495-508.
Swanson, D.A., Tayman, J., & Barr, C.F. (2000). A note on the measurement of accuracy for subnational demographic estimates. *Demography, 37*, 193-201.
Tang, Z., Almeida, C., & Fishwick, P.A. (1991). Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation, 57*, 303-310.
Tayman, J., & Swanson, D.A. (1999). On the validity of MAPE as a measure of population forecast accuracy. *Population Research and Policy Review, 18*, 299-322.
Tayman, J., Schafer, E., & Carter, L. (1998). The role of population size in the determination and prediction of population forecast errors: An evaluation using confidence intervals for subcounty areas. *Population Research and Policy Review, 17*, 1-20.
Tayman, J., Swanson, D.A., & Barr, C.F. (1999). In search of the ideal measure of accuracy for subnational demographic forecasts. *Population Research and Policy Review, 18*, 387-409.
Tschacher, W., & Ramseyer, F. (2009). Modeling psychotherapy process by time-series panel analysis (TSPA). *Psychotherapy Research, 19*, 469-481.
Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.
Valiyeva, E., Herrmann, N., Rochon, P.A., Gill, S.S., & Anderson, G.M. (2008). Effect of regulatory warnings on antipsychotic prescription rates among elderly patients with dementia: A population-based time-series analysis. *Canadian Medical Association Journal, 179*, 438-446.
Wasserman, P.D. (1989). *Neural computing: Theory and practice*. New York: Van Nostrand Reinhold.
Wilcox, R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press.