Technological Forecasting & Social Change 111 (2016) 235–243

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

A novel ranking procedure for forecasting approaches using Data Envelopment Analysis

Ali Emrouznejad [a,*], Bahman Rostami-Tabar [b], Konstantinos Petridis [c]

[a] Aston Business School, Aston University, Birmingham, B4 7ET, UK
[b] School of Strategy and Leadership, Coventry Business School, Coventry University, Priory Street, Coventry CV1 5FB, UK
[c] Department of Applied Informatics, 156 Egnatia str., 54006, Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 13 February 2016
Accepted 4 July 2016
Available online 5 August 2016

Keywords:
Forecasting
Accuracy measure
Data Envelopment Analysis

ABSTRACT

To compare the accuracy of different forecasting approaches an error measure is required. Many error measures have been proposed in the literature, however in practice there are some situations where different measures yield different decisions on forecasting approach selection and there is no agreement on which approach should be used. Generally forecasting measures represent ratios or percentages providing an overall image of how well fitted the forecasting technique is to the observations. This paper proposes a multiplicative Data Envelopment Analysis (DEA) model in order to rank several forecasting techniques. We demonstrate the proposed model by applying it to the set of yearly time series of the M3 competition. The usefulness of the proposed approach has been tested using the M3-competition where five error measures have been applied in and aggregated to a single DEA score.

1. Introduction

Measuring forecasting performance is a crucial issue. With many different methods in forecasting, understanding their relative performance is critical for more accurate prediction of the quantities of interest. Conclusions about the accuracy of various forecasting methods typically require comparisons using a range of accuracy measures. This is because different measures are designed to assess different aspects of the model. For example, Mean Square Error (MSE) puts heavier penalties on higher errors while Mean Absolute Error (MAE) is designed to lessen the effect of outliers. Various accuracy measures have been used in the literature and their properties have been discussed to some extent (Hyndman and Koehler, 2006). Obviously, it is one thing that no accuracy measure dominates the others and it is another that all reasonable accuracy measures are equally fine. Forecast accuracy evaluation becomes a more challenging task when different forecast methods/forecast scenarios and various forecast accuracy measures are involved. In a given situation, sometimes different accuracy measures will lead to different results as to which forecast method/scenario is best and they give contradictory results. These contradictory results indicate that they are not measuring the same aspect of prediction accuracy (Kitchenham et al., 2001). It has been observed through forecasting competition studies such as the M-competition (Makridakis et al., 1982) and the M3-competition (Makridakis and Hibon, 2000) that the performance of different methods changes considerably depending on the accuracy measure being used. Syntetos and Boylan (2005) stated that different accuracy measures can lead to different conclusions especially in the context of intermittent demand, where demand appears sporadically, with some time periods showing no demand at all. Chatfield (2013) argued that the best model under one criterion cannot always be the best under some other criteria. No single measure is universally best for all accuracy assessment objectives, and different accuracy measures may lead to conflicting interpretations and conclusions. Considering different forecasting approaches, we may need to produce forecasting in various forecast horizons and/or use various performance accuracy measures to assess the accuracy performance. These issues have been argued in the literature (Athanasopoulos and Hyndman, 2011; Hyndman and Koehler, 2006; Kitchenham et al., 2001; Makridakis and Hibon, 2000; Yokuma and Armstrong, 1995). However, sometimes different accuracy measures will lead to different results in terms of selecting the most accurate forecasting method. Therefore, results may be contradicting each other. Although, this problem has been encountered in the literature of forecast accuracy measurement, however no solution is proposed to facilitate the choice of best forecasting method in the condition of contradictory results. This paper is focused only on proposing a decision support system for determining the best forecasting technique based to the results of given forecasting methods and accuracy measures, rather than improving the forecasting accuracy. We are not concerned with the methods used to provide forecasts. We are interested in how applied forecasting methods can be ranked when various accuracy measures are used to evaluate the accuracy performance. Moreover, the proposed approach can also be applied in other situations such as rank different forecasting scenarios, rank forecasting methods based on the

* Corresponding author at: Aston Business School, Aston University, Birmingham, UK.
E-mail address: a.emrouznejad@aston.ac.uk (A. Emrouznejad).
URL: http://www.deazone.com (A. Emrouznejad).

http://dx.doi.org/10.1016/j.techfore.2016.07.004
0040-1625/

forecast accuracy of various horizons and single error measure. In this study Data Envelopment Analysis (DEA) methodology is used to rank the different forecasting approaches based on their values of accuracy measures. The proposed model is a multiplicative DEA model, which is mathematically shown as the right one to handle percentages or ratio data. Each forecasting technique is considered as a Decision Making Unit (DMU). Forecasting measures of each DMU are assumed to be inputs and after being log-linearised, the proposed DEA model is solved for each DMU. The forecasting techniques are ranked based on the scores obtained from DEA model (efficiency). This is an important issue from practitioner's point of view to decide which forecasting method should be selected for forecasting purposes among various approaches, especially when forecasting process is automated and hundred of thousand items need to be predicted. The results of this paper can be implemented by forecasting package software manufacturers which can add more value to their customers. The proposed multiplicative DEA model can objectively provide ranking of forecasting techniques based on efficiency scores. In the presence of ties from the ranking, three meta-frontier techniques are presented, namely cross efficiency, super efficiency and lambda frequency. Section 2 of this paper studies the background of Data Envelopment Analysis and forecast comparison. Section 3 describes the proposed DEA model to select the best forecasting approach. In Section 4 an application of the proposed method on yearly M3-competition time series and forecasting methods is demonstrated and the results are discussed. Conclusions are drawn in Section 5.

2. Background and related works

2.1. Introduction to Data Envelopment Analysis

Data Envelopment Analysis (DEA), is a method for assessing the comparative performance of units (DMUs) converting a set of inputs to a bundle of outputs, based on certain assumptions. The first models of DEA technique have been proposed by Charnes et al. (1978) and Banker et al. (1984). Thereafter, the area of DEA has been largely expanded with extensions to the aforementioned works. The main characteristic of the DEA technique is its ability to provide a unified efficiency score of an assumed production process where inputs are consumed in order to produce outputs, which in most cases are desirable, though undesirable outputs may occur as well (Seiford and Zhu, 2002). For further details about DEA and its application see Emrouznejad and De Witte (2010) and Cook and Seiford (2009). In cases where the weights of the model provide zero values, then a different multiplicative DEA model must be used.

2.2. Comparison of forecasting techniques

Forecasting is designed to help decision making and planning in the present by predicting possible future alternatives. In the taxonomy of forecasting methods (Yokuma and Armstrong, 1995), judgmental and statistical forecasting are the two main categories (Hyndman and Athanasopoulos, 2014). The assessment of forecasting techniques is an interesting subject that has been addressed throughout the years (De Gooijer and Hyndman, 2006). Some studies compared accuracy measured with other criteria such as ease of use, ease of interpretation, cost saving, etc. on forecast evaluation. They concluded that accuracy was the most important criterion for evaluating forecasting techniques (Collopy and Armstrong, 1992; Witt and Witt, 1992). In most of the cases, forecasting techniques are compared against the values of accuracy measures or are examined by situation or data used. The accuracy measures that are often used in order to evaluate the quality of a forecasting technique are Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Based on the study of Collopy and Armstrong (1992) that has been conducted with a panel of 49 experts in the field of forecasting, 85% of the respondents consider accuracy measures from important (56%) to extremely important (29%). Based on the aforementioned study, several works have been published assessing quantitative and qualitative criteria of forecasting techniques (Yokuma and Armstrong, 1995). In that multi-aspect study, an agreement analysis has been performed using a questionnaire survey measuring the opinions of 322 experts divided into 4 categories, namely Decision Maker (DM), Practitioner, Educator and Researcher. Among the questions asked, the largest average agreement score was that of "Accuracy" whereas "Timeliness" in providing forecasts' gathered the second largest score. Comparison of forecasting techniques can also be conducted by testing the techniques, applying to real life data sets representing sales (Abdel-Khalik and El-Sheshai, 1983; Geurts and Patrick Kelly, 1986). It is important to note that due to the advantages and disadvantages of each accuracy measure, no single error measure can capture all aspects of accuracy. Many forecast accuracy measures have been proposed in the literature and the recommendations for selecting the appropriate error measurements are discussed. Authors argued that generally utilization of various accuracy measures should be more efficient, as each accuracy measure may look at a different aspect of accuracy (Kitchenham et al., 2001). A summary of some of the issues is given by (Davydenko and Fildes, 2013; De Gooijer and Hyndman, 2006; Fildes et al., 2011; Hyndman and Koehler, 2006) and (Yokuma and Armstrong, 1995). De Gooijer and Hyndman (2006) reviewed a variety of accuracy measures used in the literature to evaluate the performance of forecasting methods up to 2005. Hyndman and Koehler (2006) provided a critical survey on various accuracy measures. Fildes et al. (2011) argued that no single error measure captures the distributional features of the errors when summarized across data series and discussed four error measures that should capture the essential characteristics of the forecast results. Davydenko and Fildes (2013) discussed many error measures by focusing on the performance measurement of judgemental forecasting.

2.3. DEA score as a means for selecting best forecasting techniques

To rank forecasting techniques several approaches have been introduced in the literature using Machine Learning, Data Mining techniques and forecasting with Neural Networks based on their measures. For example, using Machine Learning techniques several indices have been developed such as Adjusted Ratio of Ratios (ARR) which is a multi-criteria evaluation index and resembles to the efficiency measure as it provides the relative efficiency of each technique (Brazdil et al., 2003). However, the performance is relative and concerns only the comparison of two algorithms given a compromise (trade-off) between two criteria. Another shortfall of the proposed approach is that in order to extend the comparisons to more than two algorithms, certain aggregations of the criteria must be made. On the contrary, the efficiency score from DEA technique is objectively extracted. The efficiency of the calculation process using Data Mining techniques is similar to Machine learning. Efficiency is formed as a fraction of the weighted outputs to inputs for of each technique (Nakhaeizadeh and Schnabl, 1997; Nakhaeizadeh and Schnabl, 1998). In a different context, technological forecasting has been examined by Lim et al. (2014). Based on this method, the technological capabilities of different technologies are assessed based on DEA. Technological Forecasting using Data Envelopment Analysis (TFDEA) has also been applied on fighter jet and commercial technology by Inman et al. (2006). However, the proposed DEA method, assesses different versions of DMUs based on inputs and outputs and does not provide a decision support system for determining the best forecasting technique. Later studies prove that classical DEA models are not appropriate to handle percentage or ratio data (Emrouznejad et al., 2010). Duong (1988) mentioned that it is not uncommon in practice to have a set of forecasts which yield different rankings of the underlying techniques for different performance criteria. A hierarchical approach to rank the forecasting techniques has been suggested using the Analytic Hierarchy Process (AHP) as a general framework for obtaining the weights for forecasts combination. Pairwise comparisons of forecast