

A novel ranking procedure for forecasting approaches using Data Envelopment Analysis

Ali Emrouznejad

Aston Business School, Aston University, Birmingham B4 7ET, UK

Bahman Rostami-Tabar

School of Strategy and Leadership, Coventry Business School, Coventry University, Priory Street, Coventry, CV1 5FB, UK

Konstantinos Petridis

Department of Applied Informatics, 156 Egnatia str., 54006, Thessaloniki, Greece

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.

Keywords: Forecasting, Accuracy measure, Data Envelopment Analysis

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)

Email address: a.emrouznejad@aston.ac.uk (Ali Emrouznejad)

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 [1]. 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 [2]. It has been observed through forecasting competition studies such as the M-competition [3] and the M3-competition [4] that the performance of different methods changes considerably depending on the accuracy measure being used. Syntetos and Boylan (2005) [5] 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) [6] 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 ([1], [2], [4], [7], [8]). 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 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, Cooper, & Rhodes (1978) [9] and Banker, Charnes, & Cooper 1984 [10]. 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 [11]. For further details about DEA and its application see Emrouznejad & De Witte (2010) [12] and Cook et al., 2009 [13]. 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 [7] judgmental and statistical forecasting are the two main categories [14]. The assessment of forecasting techniques is an interesting subject that has been addressed throughout the years [15]. 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 ([16], [17]). 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) [16] 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 [7]. 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 ([18], [19]). 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 [2]. A summary of some of the issues is given by [1], [15],[20], [21] and [22]. De Gooijer & Hyndman(2006) [15] reviewed a variety of accuracy measures used in the literature to evaluate the performance of forecasting methods up to 2005. Hyndman and Koehler (2006)[1] provided a critical survey on various accuracy measures. Field et al. (2011) [20] 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 & Fildes (2013) [21] 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 one of which is 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 [23]. 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 ([24], [25]). In a different context, technological forecasting has been examined by Lim, et al. (2014) [26]. 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) [27]. 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 [28]. Duong (1988) [29] 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 measures have been determined from a panel of experts provided weights leading to the ranking of each technique. The ranking result is purely based on subjective comparisons of the groups of experts, thus different group of experts could provide different comparisons and consequently, different forecasting methods ranking. Therefore, the ranking result is not robust based because of the discussed shortfall of AHP. Standard DEA has also been applied also for the evaluation of forecasting of Neural Networks [30]. Similar to the evaluation of the aforementioned Data Mining and Machine Learning techniques presented above, efficiency is derived as weighted fraction of outputs to inputs. Let's assume that there are n DMUs ($DMU_j; j = 1, 2, \dots, n$), consuming m inputs ($x_{i,j}; i = 1, 2, \dots, m$), and producing s outputs ($y_{r,j}; r = 1, \dots, s$). Two main efficiency models in DEA are input-oriented and out-put oriented, as formulated in Model 1 and Model 2.

In this model, λ_j provides information for the reference set (peers) in the case where DMU under investigation (j_o) is not efficient ($\theta_{j_o} \neq 1$ or $\phi_{j_o} \neq 1$). Models 1 and 2 are the first DEA models introduced by Charnes, Cooper, & Rhodes (1978) [9]. DEA has been used to evaluate forecasting techniques including ARIMA (Autoregressive Integrated Moving Average), RW (Random Walk), Vector Error Correction (VEC), Regression and Error Correction models [31]. Furthermore, basic DEA Models 1 and 2 along with Super-Efficiency models have been applied to the ranking of 14 chosen competing volatility forecasting approaches, applying crude oil price data [31]. However, as we will show applying standard DEA models to forecast accuracy measures may yield incorrect results. Accuracy measures represent ratios thus standard DEA models provide false results due to loss of information while the returns to scale orientation (RTS) is by default Constant Returns to Scale (CRS) [32]. Emrouznejad et al.(2010) [28] have shown that in the case of ratios in inputs and outputs, multiplicative DEA formulations are more appropriate as the concept of the geometric mean with non-dimensional (unit invariance) properties is used. Furthermore, it is shown in the literature that in the case where data are ratios, classical DEA models provides wrong results as the Production Possibility Set (PPS) is not additive, but multiplicative. Denoting PPS as $P = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \mapsto \mathbf{y}\}$, assuming that inputs x can produce outputs y (whereas x and y represent vectors of inputs and outputs, respectively), then DEA provides an estimation of the PPS. Based on the convexity axiom of DEA ([10]), a convex combination of any two

Model 1: VRS input orientated standard DEA

$$\begin{aligned} & \min \theta_{j_o} \\ & \text{s.t.} \\ & \sum_{j=1}^n \lambda_j \cdot x_{i,j} \leq \theta_{j_o} \cdot x_{i,j_o}, i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j \cdot y_{r,j} \geq y_{r,j_o}, r = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0, j = 1, \dots, n \\ & \theta \text{ free} \end{aligned}$$

Model 2: VRS output orientated standard DEA

$$\begin{aligned} & \max \phi_{j_o} \\ & \text{s.t.} \\ & \sum_{j=1}^n \lambda_j \cdot x_{i,j} \leq x_{i,j_o}, i = 1, 2, \dots, m \\ & \sum_{j=1}^n \lambda_j \cdot y_{r,j} \geq \phi_{j_o} \cdot y_{r,j_o}, r = 1, 2, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0, j = 1, \dots, n \\ & \phi \text{ free} \end{aligned}$$

points on the PPS also belongs to PPS. In the case a single input, two outputs, representing ratios, the PPS is graphically represented in Figure 1. If P_A , P_B are two points of PPS such that $P_A(x_A, y_A)$ and $P_B(x_B, y_B)$, then assuming that point P' is a convex combination of P_A and P_B , then $P' = \lambda \cdot P_A + (1 - \lambda) \cdot P_B$. Taking values for λ , the new point resulting from the convex combination of P_A and P_B , does not belong to the PPS. A relevant example and proof of the aforementioned procedure is demonstrated in Emrouznejad & Amin (2009) [33]. Hence, the model that can be applied in this case is a multiplicative DEA model as developed by Emrouznejad et al. (2010) [28].

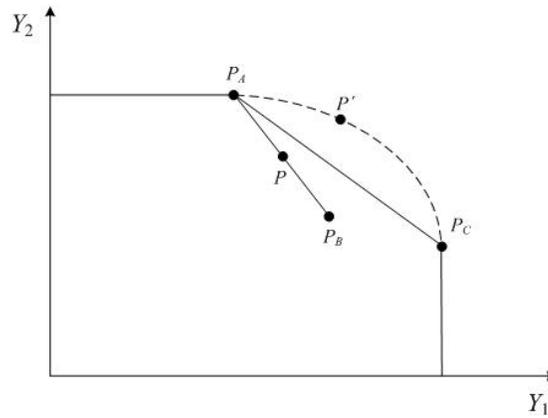


Figure 1: An output Productivity Possibility Set

3. Proposed approach

3.1. Multiplicative DEA model

Following the discussion in Section 2.3, the multiplicative DEA model that is used to rank the forecasting techniques applied to yearly time series M3-competition data [4], is the following [35]:

As described in the Model 3, the proposed multiplicative DEA model is input oriented, considering geometric convexity as any point of the new PPS can be calculated based on the relation $P'' = (P_A)^\lambda \cdot P_B^{1-\lambda}$. In Model 3, $x_{i,j}$ denotes i input of DMU j whereas λ_j denotes the coefficient to peers of each DMU j ; θ_{j_o} is the efficiency score for each DMU under investigation o . The above Model 3 is non-linear due to existence of exponential terms (α^x). In order to get an optimal solution, an equivalent model is proposed based on log-linearization of Model 3. For log-linearization, the following \log_{10} function is used:

Model 3: A multiplicative input-oriented DEA model

$$\begin{aligned}
 & \min \theta_{j_o} \\
 & \text{s.t.} \\
 & \prod_{j=1}^n (x_{i,j})^{\lambda_j} \leq \theta_{j_o} \cdot x_{i,j_o}, i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j \geq 0, j = 1, \dots, n \\
 & \theta \text{ free}
 \end{aligned}$$

Model 4: Log linearisation of multiplicative input-oriented DEA model

$$\begin{aligned}
 & \min \tilde{\theta}_{j_o} \\
 & \text{s.t.} \\
 & \sum_{j=1}^n \lambda_j \cdot \tilde{x}_{i,j} \leq \tilde{x}_{i,j_o} + \tilde{\theta}_{i,j_o}, i = 1, 2, \dots, m \\
 & \sum_{j=1}^n \lambda_j = 1 \\
 & \lambda_j \geq 0, j = 1, \dots, n \\
 & \tilde{\theta} \text{ free}
 \end{aligned}$$

$$\begin{aligned}
 & \log_{10} \left(\prod_{j=1}^n (x_{i,j})^{\lambda_j} \right) \leq \log_{10}(x_{i,j_o}) \Leftrightarrow \\
 & \left(\sum_{j=1}^n \log_{10}(x_{i,j}) \cdot \lambda_j \right) \leq \log_{10}(x_{i,j_o}) + \log_{10}(\theta_{i,j_o}), i = 1, 2, \dots, m, j = 1, \dots, n
 \end{aligned}$$

Taking the logarithm on both sides of the first constraint of Model 4 is derived.

In Linear Programming (LP) Model 4, $\tilde{x}_{i,j}$ denotes the $\log_{10}(x_{i,j})$. Objective function seeks the minimization of log-efficiency of each DMU o ($\tilde{\theta}_{j_o}$) and efficiency scores are obtained by calculating $10^{\tilde{\theta}_{j_o}^*}$ for each DMU, where $\tilde{\theta}_{j_o}^*$ is the optimal solution from Model 4.

It must be noted that as Model 4 may handle accuracy data per DMU less than one, then logarithm yields a negative value. This is resolved with

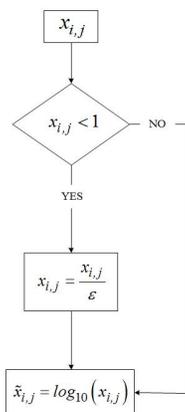


Figure 2: Normalisation process for process for input data

normalization procedure, as stated in 2. Given the initial data (input i for each DMU j), if all the data for inputs of all DMUs are above 1, then the new logarithmic input vectors are computed. However, for an input less than 1, then the corresponding column is divided by the minimum number normalizing data to avoiding negative values after logarithm calculation.

3.2. Finding the best forecasting technique in case of ties

When using DEA to rank forecasting techniques, it is quite common to get ties of ranking due to the fact that two or more forecasting techniques become efficient. In order to deal with those cases where there is a tie in the ranking of forecasting techniques, cross efficiency, super efficiency and lambda frequency can be used.

Cross efficiency : The idea of cross efficiency [36] is to support the weak discrimination of basic DEA models. In order to compute the cross efficiency the following steps are followed. Firstly, optimal solutions for weights variable (in the dual of Model 4) are derived. Then the cross efficiency of each DMU under investigation is calculated based on the weights of other DMUs. For full description of cross efficiency see [36].

Super efficiency: Super efficiency ([37], [38]) has been widely used in order to assess the efficiency of DMUs and to provide more discrimination power to basic DEA models. The concept of super efficiency is based on assessing the efficiency of all the DMUs while excluding the DMU under investigation from the left hand side of Model 4, this is presented in Model 5. The proposed model with the concept of super efficiency is presented in Model 8.

Model 5: Super Efficiency DEA model

$$\begin{aligned}
 & \min \tilde{\theta}_{j_o} \\
 & \text{s.t.} \\
 & \sum_{\substack{j=1 \\ j \neq j_o}}^n \lambda_j \cdot \tilde{x}_{i,j} \leq \tilde{x}_{i,j_o} + \tilde{\theta}_{i,j_o}, i = 1, 2, \dots, m \\
 & \sum_{\substack{j=1 \\ j \neq j_o}}^n \lambda_j = 1 \\
 & \lambda_j \geq 0, j = 1, \dots, n \\
 & \tilde{\theta} \text{ free}
 \end{aligned}$$

Lambda frequency: When solving a DEA problem, besides the results that concern efficiency, lambda values (peers) are also calculated. These values are interpreted as the resemblance of the DMU under investigation with the units that belong to its reference set. Non-efficient units have as peers (positive lambda values) only efficient DMsUs. Thus, in order to discriminate the ranking in case of a tie between two or more DMUs then counting only the instances where an efficient DMU is a peer in the reference set of a non-efficient DMU, can be used as a discriminating measure.

3.3. An illustrative example

In order to provide better understanding of the proposed approach and to make it reproducible, an illustrative example is presented with 5 forecasting techniques. For each forecasting technique, 3 error measures have been calculated to evaluate its accuracy. From this figure, we can not clearly answer this question, “which forecasting technique is the best?”. The optimal point would be the one that would minimize all accuracy measures simultaneously. Let us look at *FOR02* and *FOR05* in Figure 3; it can be seen that *FOR02* performs better in measure 1 comparing with measures 2 and 3. On the contrary, *FOR05* performs better in measures 2, 3 and worst in measure 3. The data for each forecasting technique are presented in Table 1.

As indicated in Table 1, the values of measure M_1 for each forecasting technique are greater than 1 and the \log_{10} can be calculated straightforward. The same stands for measure M_2 but as can be seen in the last measure (M_3), the value of forecasting technique 1 (*FOR01*) is less than 1. Thus, calculating the \log_{10} value would provide negative results, thus based on log-normalisation procedure (Figure 2), all the elements of the column M_3 are divided by a small positive infinitesimal number ($\epsilon = 10^{-5}$). After that normalization, the \log_{10} value of the last column is calculated as well. The results are presented in Table 2.

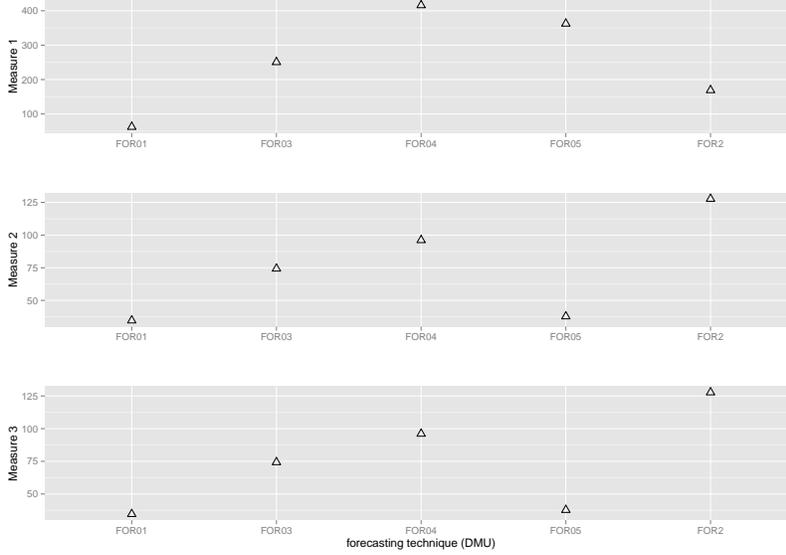


Figure 3: Comparison of 5 forecasting techniques based on their accuracy measures

Table 1: Data for illustrative example

Forecasting Technique	M_1	M_2	M_3
<i>FOR01</i>	62.24	37.618	0.0377
<i>FOR02</i>	168.937	127.708	42.479
<i>FOR03</i>	250.522	74.292	22.614
<i>FOR04</i>	415.936	96.076	7.063
<i>FOR05</i>	362.426	34.747	4.741

Based on Model 4, the following LP models are solved for each DMU (*FOR01*, ..., *FOR05*). By solving the LP models as presented in Table 3, the optimal values regarding $\tilde{\theta}^*$, $\theta^* = 10\tilde{\theta}^*$ and λ^* variables are derived.

- *FOR01*: $\tilde{\theta}^* = 0$, $\theta^* = 1$ (100%), ($\lambda_1^* = 1$)
- *FOR02*: $\tilde{\theta}^* = -0.434$, $\theta^* = 0.368$ (36.8%) ($\lambda_1^* = 1$)
- *FOR03*: $\tilde{\theta}^* = -0.309$, $\theta^* = 0.491$ (49.1%), ($\lambda_1^* = 0.613$, $\lambda_5^* = 0.387$)
- *FOR04*: $\tilde{\theta}^* = -0.425$, $\theta^* = 0.375$ (37.5%) ($\lambda_1^* = 0.478$, $\lambda_5^* = 0.522$)
- *FOR05*: $\tilde{\theta}^* = 0$, $\theta^* = 1$ (100%), ($\lambda_5^* = 1$)

Forecasting techniques are ranked based on the efficiency score. *FOR01* and *FOR05* are ranked first, *FOR03* is ranked second, *FOR04* is ranked third

Table 2: Log values of data of illustrative example

Forecasting Technique	$\log_{10}(M_1)$	$\log_{10}(M_2)$	$\log_{10}\left(\frac{M_3}{\epsilon}\right)$
<i>FOR01</i>	1.794	1.575	3.576
<i>FOR02</i>	2.228	2.106	6.628
<i>FOR03</i>	2.399	1.871	6.354
<i>FOR04</i>	2.619	1.983	5.849
<i>FOR05</i>	2.559	1.541	5.676

and *FOR02* is ranked fourth. As it can be seen, there is a tie in the ranking between *FOR01* and *FOR05*. Based on lambda frequency, presented in Section 3.2, the count of positive λ^* values that are referred to *FOR01* are 3 while the corresponding number for *FOR05* are 2. Accordingly, we can conclude that the rankings are as follows: *FOR01* \succ *FOR05* \succ *FOR03* \succ *FOR04* \succ *FOR02*.

Table 3: LP formulations for each forecasting technique of the illustrative example

DMU	LP formulation
FOR01	$\begin{aligned} & \min \tilde{\theta} \\ & \text{s.t.} \\ & 1.794\lambda_1 + 2.228\lambda_2 + 2.399\lambda_3 + 2.619\lambda_4 + 2.559\lambda_5 - \tilde{\theta} \leq 1.794 \\ & 1.575\lambda_1 + 2.106\lambda_2 + 1.871\lambda_3 + 1.983\lambda_4 + 1.541\lambda_5 - \tilde{\theta} \leq 1.575 \\ & 3.576\lambda_1 + 6.628\lambda_2 + 6.354\lambda_3 + 5.849\lambda_4 + 5.676\lambda_5 - \tilde{\theta} \leq 3.576 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0, \tilde{\theta} \text{ free} \end{aligned}$
FOR02	$\begin{aligned} & \min \tilde{\theta} \\ & \text{s.t.} \\ & 1.794\lambda_1 + 2.228\lambda_2 + 2.399\lambda_3 + 2.619\lambda_4 + 2.559\lambda_5 - \tilde{\theta} \leq 2.228 \\ & 1.575\lambda_1 + 2.106\lambda_2 + 1.871\lambda_3 + 1.983\lambda_4 + 1.541\lambda_5 - \tilde{\theta} \leq 2.106 \\ & 3.576\lambda_1 + 6.628\lambda_2 + 6.354\lambda_3 + 5.849\lambda_4 + 5.676\lambda_5 - \tilde{\theta} \leq 6.628 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0, \tilde{\theta} \text{ free} \end{aligned}$
FOR03	$\begin{aligned} & \min \tilde{\theta} \\ & \text{s.t.} \\ & 1.794\lambda_1 + 2.228\lambda_2 + 2.399\lambda_3 + 2.619\lambda_4 + 2.559\lambda_5 - \tilde{\theta} \leq 2.399 \\ & 1.575\lambda_1 + 2.106\lambda_2 + 1.871\lambda_3 + 1.983\lambda_4 + 1.541\lambda_5 - \tilde{\theta} \leq 1.871 \\ & 3.576\lambda_1 + 6.628\lambda_2 + 6.354\lambda_3 + 5.849\lambda_4 + 5.676\lambda_5 - \tilde{\theta} \leq 6.354 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0, \tilde{\theta} \text{ free} \end{aligned}$
FOR04	$\begin{aligned} & \min \tilde{\theta} \\ & \text{s.t.} \\ & 1.794\lambda_1 + 2.228\lambda_2 + 2.399\lambda_3 + 2.619\lambda_4 + 2.559\lambda_5 - \tilde{\theta} \leq 2.619 \\ & 1.575\lambda_1 + 2.106\lambda_2 + 1.871\lambda_3 + 1.983\lambda_4 + 1.541\lambda_5 - \tilde{\theta} \leq 1.983 \\ & 3.576\lambda_1 + 6.628\lambda_2 + 6.354\lambda_3 + 5.849\lambda_4 + 5.676\lambda_5 - \tilde{\theta} \leq 5.849 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0, \tilde{\theta} \text{ free} \end{aligned}$
FOR05	$\begin{aligned} & \min \tilde{\theta} \\ & \text{s.t.} \\ & 1.794\lambda_1 + 2.228\lambda_2 + 2.399\lambda_3 + 2.619\lambda_4 + 2.559\lambda_5 - \tilde{\theta} \leq 2.559 \\ & 1.575\lambda_1 + 2.106\lambda_2 + 1.871\lambda_3 + 1.983\lambda_4 + 1.541\lambda_5 - \tilde{\theta} \leq 1.541 \\ & 3.576\lambda_1 + 6.628\lambda_2 + 6.354\lambda_3 + 5.849\lambda_4 + 5.676\lambda_5 - \tilde{\theta} \leq 5.676 \\ & \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\ & \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \geq 0, \tilde{\theta} \text{ free} \end{aligned}$

4. A real application to M3-Competition data

We demonstrate use of the proposed approach by applying it in the yearly M3–competition data [4]. We consider 22 forecasting methods discussed in the M3– competition. Our aim is to observe the data at each forecast period and calculate the forecast accuracy assuming that we have different forecast approaches and various forecast performance measures. Please refer to Makridakis and Hibon (2000) [4] for details of forecasting methods used in this study. In Table 4 the forecasting techniques used in this instance are demonstrated and are presented as DMUs of the proposed DEA model.

Table 4: Forecasting techniques used in M3-Competition time series data and in DEA analysis

Forecasting Technique	DMU	Forecasting Technique	DMU
ROBUST	FOR01	PP Autocast	FOR12
AutoBox2	FOR02	ForecastPro	FOR13
ForcX	FOR03	Flors.Pearsc1	FOR14
RBF	FOR04	SMARTF	FOR15
SINGLE	FOR05	B.J.aut	FOR16
THETA _{sm}	FOR06	AutoBox3	FOR17
NAIVE2	FOR07	DAMPEN	FOR18
THETA	FOR08	ARARMA	FOR19
Auto.ANN	FOR09	WINTER	FOR20
Flors.Pearc2	FOR10	HOLT	FOR21
COMB.S.H.D	FOR11	AutoBox1	FOR22

4.1. Forecasting performance evaluation

The measures used for comparing the forecasting accuracy are the root mean square error (RMSE), Mean Absolute deviation (MAE), mean absolute percentage error (MAPE), Symmetric Mean Absolute Percentage Error (sMAPE) and Mean Absolute Scaled Error (MASE) which are defined as follows (Hyndman and Koehler, 2006):

1. Root Mean Square Error (RMSE) = $\sqrt{\frac{1}{n} \sum_{i=1}^n e_t^2}$
2. Mean Absolute Error (MAE) = $\frac{1}{n} \sum_{i=1}^n |e_t|$
3. Mean Absolute Percentage Error (MAPE) = $\frac{1}{n} \sum_{i=1}^n |100 \cdot \frac{e_t}{Y_t}|$
4. Symmetric Mean Absolute Percentage Error (sMAPE) = $\frac{1}{n} \sum_{i=1}^n \frac{100 \cdot |e_t|}{Y_t + F}$
5. Mean Absolute Scaled Error (MASE) = $\frac{1}{n} \sum_{t=1}^n \left(\frac{e_t}{\frac{1}{n-1} \cdot \sum_{i=2}^n |Y_i - Y_{i-1}|} \right)$

4.2. Case Study

The case study that is selected to demonstrate the applicability of the presented approach, uses times series data retrieved from M3–Competition, where

Model 6: Linearization of multiplicative input-oriented DEA model

$$\begin{aligned}
& \min \tilde{\theta}_{j_o} \\
& \text{s.t.} \\
& \sum_{j=1}^{22} \lambda_j \cdot \widetilde{RMSE}_{i,j} \leq \widetilde{RMSE}_{i,j_o} + \tilde{\theta}_{j_o} \\
& \sum_{j=1}^{22} \lambda_j \cdot \widetilde{MAE}_{i,j} \leq \widetilde{MAE}_{i,j_o} + \tilde{\theta}_{j_o} \\
& \sum_{j=1}^{22} \lambda_j \cdot \widetilde{MAPE}_{i,j} \leq \widetilde{MAPE}_{i,j_o} + \tilde{\theta}_{j_o} \\
& \sum_{j=1}^{22} \lambda_j \cdot \widetilde{sMAPE}_{i,j} \leq \widetilde{sMAPE}_{i,j_o} + \tilde{\theta}_{j_o} \\
& \sum_{j=1}^{22} \lambda_j \cdot \widetilde{MASE}_{i,j} \leq \widetilde{MASE}_{i,j_o} + \tilde{\theta}_{j_o} \\
& \sum_{j=1}^{22} \lambda_j = 1 \\
& \lambda_j \geq 0, j = 1, \dots, 22 \\
& \tilde{\theta} \text{ free}
\end{aligned}$$

the evaluation of efficiency has been performed based on five measures, namely RMSE, MAE, MAPE, sMAPE and MASE, derived from 22 forecasting techniques. In this section, the application of the proposed DEA approach to yearly M3-Competition data set is demonstrated. As shown in Figure 4, 22 forecasting techniques are applied providing values for five measures (RMSE, MAE, MAPE, sMAPE and MASE). As it can be seen, more or less all forecasting techniques provide a uniform image with slight modifications per measure. The following DEA model is applied to extract the efficiency and rank DMUs based on the given measures. In Model 6 $(\tilde{\cdot})$ express the $\log_{10}(\cdot)$ and a VRS technology is assumed as well.

Figure 4 shows the original error measures of forecasting techniques while Figure 5 presents these measures after a logarithmic transformation. Due to the fact that logarithmic function is monotonically increasing function, the trend is the same between actual data and this representation. The last graph of Figure 5 yields the efficiency of each forecasting technique in the range of $[0, 1]$ after applying the transformation $10^{\tilde{\theta}}$. The results are also presented in Table 5. Solving the dual of Model 4, the weights reported are not zero.

In Table 5 the efficiency score is presented for each forecasting technique.

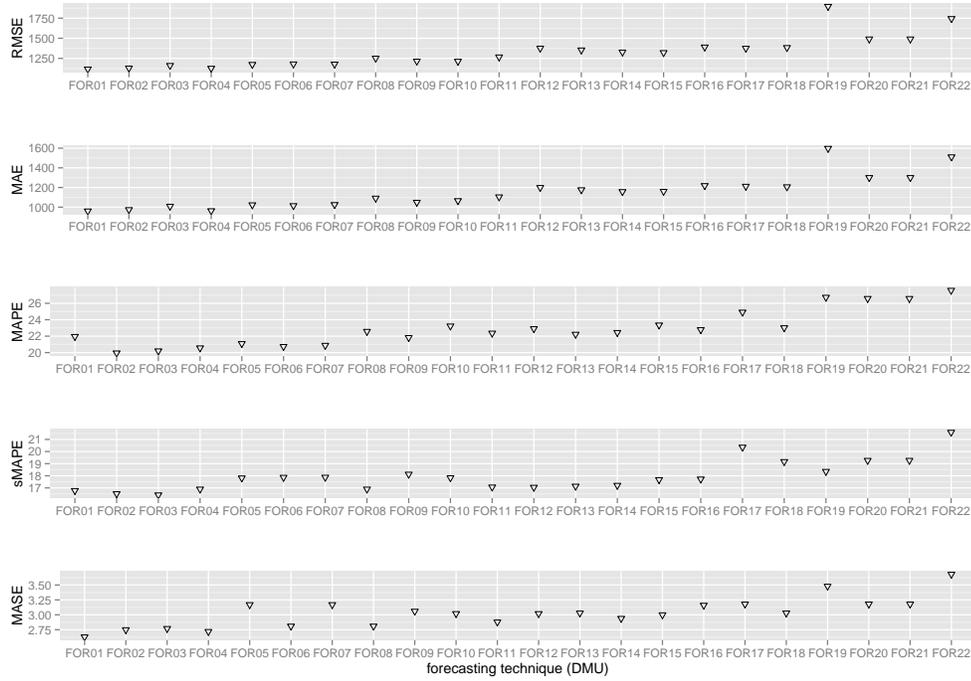


Figure 4: Comparison of the 22 forecasting techniques applied to time series data (M3- Competition)

Table 5: Efficiency and ranking of forecasting techniques based on the proposed DEA model (M3 Competition data)

DMU	DEA score	Rank	DMU	DEA score	Rank
<i>FOR01</i>	1	1	<i>FOR14</i>	0.954096	12
<i>FOR02</i>	1	1	<i>FOR15</i>	0.928733	13
<i>FOR03</i>	1	1	<i>FOR10</i>	0.928525	14
<i>FOR04</i>	1	1	<i>FOR09</i>	0.927467	15
<i>FOR08</i>	0.975753823	5	<i>FOR16</i>	0.926114	16
<i>FOR06</i>	0.973362993	6	<i>FOR18</i>	0.894389	17
<i>FOR12</i>	0.963049853	7	<i>FOR19</i>	0.894336	18
<i>FOR11</i>	0.9619215	8	<i>FOR20</i>	0.857651	19
<i>FOR05</i>	0.95990411	9	<i>FOR21</i>	0.857651	20
<i>FOR13</i>	0.957992999	10	<i>FOR17</i>	0.843845	21
<i>FOR07</i>	0.957929481	11	<i>FOR22</i>	0.760537	22

Based on the data from yearly *M3*–Competition, forecasting techniques *FOR01*, *FOR02*, *FOR03* and *FOR04* have the largest efficiency and are ranked as 1. The extracted efficiency is used as a unified measure based on which forecasting techniques

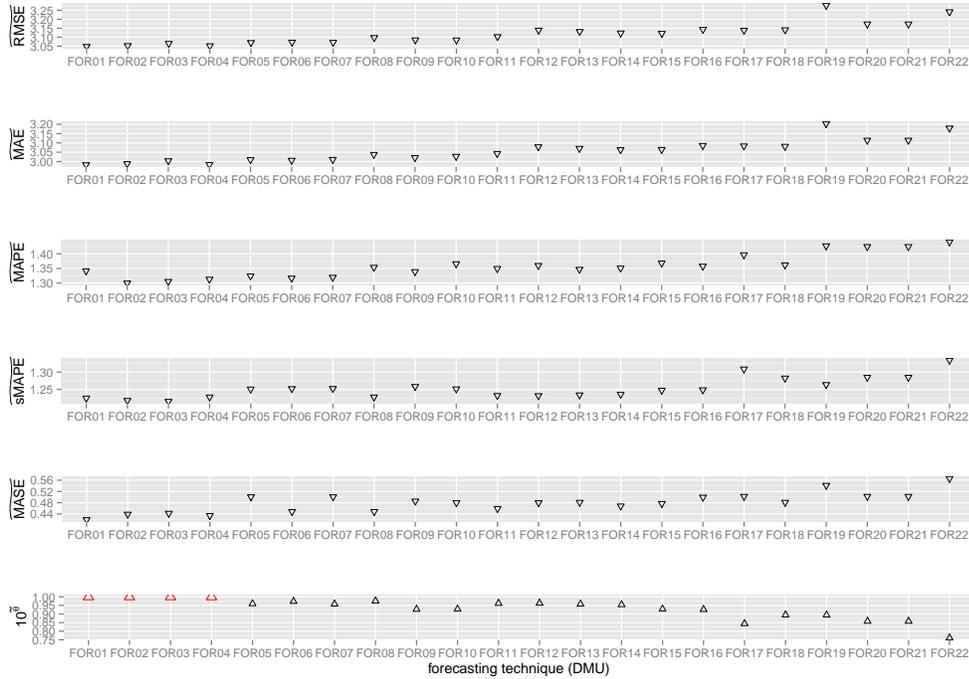


Figure 5: Logarithmic transformation of accuracy measures and efficiency (sales from M3-Competition)

are ranked. The importance of the proposed approach is that the efficiency is extracted after comparing the values of measures of each DMU against each other and the score is calculated objectively. This technique is served as a multi-criteria decision making technique, as each forecasting technique is not performing the same in all measures. Moreover, as each proposed measure is equally important to evaluation of each forecasting technique, the score provided can be used for assessing forecasting techniques based on one or more measures. One of the disadvantages of DEA is that by construction, the technique is designed to measure efficiency/productivity and not providing a single winner; in this case ranking of forecasting techniques is performed upon multiplying the values of different criteria.

4.3. Handling ties

Another way to determine the true ranking of forecasting techniques in case of a tie is to measure the lambda frequency. More specifically, by measuring how many times the non-efficient DMUs refer to them in their reference set, then the higher the frequency, the better for that forecasting technique (FOR). Taking as an example the results for lambda values from Table 6, then the frequency of positive lambda values are shown in the last row which it can be

Table 6: Peers and Lambda frequency of each forecasting technique

DMU	λ_1	λ_2	λ_3	λ_4
<i>FOR01</i>	1			
<i>FOR02</i>		1		
<i>FOR03</i>			1	
<i>FOR04</i>				1
<i>FOR05</i>	0.15311	0.846889		
<i>FOR06</i>	0.12135	0.878642		
<i>FOR07</i>	0.026352	0.973648		
<i>FOR08</i>	0.19682		0.803818	
<i>FOR09</i>	0.153105	0.846895		
<i>FOR10</i>	0.173738	0.826262		
<i>FOR11</i>			1	
<i>FOR12</i>			1	
<i>FOR13</i>			1	
<i>FOR14</i>			1	
<i>FOR15</i>			1	
<i>FOR16</i>			1	
<i>FOR17</i>	0.54925	0.45075		
<i>FOR18</i>	0.328412	0.671588		
<i>FOR19</i>			1	
<i>FOR20</i>	0.299323		0.700677	
<i>FOR21</i>	0.299323		0.700677	
<i>FOR22</i>			1	
Number of positive λ	11	8	12	1

seen that the DMU with the highest lambda frequency is *FOR03*, whereas the forecasting technique with the lowest lambda frequency is *FOR04*. The ranking obtained based on Table 6 is $FOR03 \succ FOR01 \succ FOR02 \succ FOR04$.

5. Overview of the proposed approach

In this paper a DEA procedure for ranking forecasting techniques has been proposed. Based on this procedure, each forecasting technique is treated as a DMU and forecasting errors of each technique as inputs. As forecasting measures are ratio data, the proposed DEA model is multiplicative. In order to handle the data, a log-normalisation of the data is performed in advance (Figure 2). The use of a multiplicative DEA model in this paper, against classical DEA models, is mathematically justified in Section 3. Introducing the log-normalised data to Model 4, forecasting techniques are ranked based on their efficiency. In the presence of more than one "winner", cross efficiency, super efficiency or Lambda frequency approaches can be used as meta-frontier analysis techniques. The strength of the proposed model is its capability to provide an objective ranking of forecasting techniques, independently of the number of measures involved. The measures are all taken into consideration and a multiplicative DEA model is used as a benchmarking technique. Ranking derived from the proposed technique is robust as it is based on the results of efficiency score. Robustness of the ranking proposed based on multiplicative DEA model is also justified due to the absence of a Decision Maker; unlike AHP, Analytic Network Process (ANP) or MCDM techniques where DMs provide pairwise comparisons (or weights to criteria) for forecast measures, multiplicative DEA leaves no room for DM interventions. However, if necessary, extra constraints can be introduced to model preference (weights) to each forecasting measure compared to others. Another strength of the proposed approach is its simplicity. The user only provides the data (forecasting measures), while the ranking is derived by solving iteratively the multiplicative DEA model for each DMU. Despite its strengths, the proposed approach has also few weaknesses. As DEA is a benchmarking technique, it is quite common for two or more DMUs to be efficient, leading to ties in the corresponding ranking. This shortfall is handled by applying cross efficiency, super efficiency or lambda frequency techniques to DMUs that are efficient and have ties in the ranking.

6. Conclusions

The discipline of forecasting is very important as it has grown over the last years given the number of articles that have been published. The importance of forecasting is the development of techniques that predict future values for demand, consumption of energy, energy, oil prices etc. However, the acceptance or rejection of a technique is based on how well the model is fitted to the data set, for in-sample adjustments and afterwards for out-of-sample forecasting. For that reason, forecasting or accuracy measures have been proposed to check the performance of a forecasting technique when applied to a time series data set. Nevertheless, as there is a plethora of accuracy measures the forecasting technique may perform better in one measure but not in another measure comparing to another forecasting technique. In order to overcome this obstacle, a model that provides an overall score that takes all the different accuracy measures into account and forecasting techniques will be ranked upon, is proposed. The proposed multiplicative DEA model is used as the data (accuracy measures) express percentages and in this case, additive models that have been used to provide a unified score of forecasting techniques provide wrong results and do not satisfy fundamental principles of DEA. The data are log-normalised and the multiplicative DEA model is solved for each DMU. In the presence of ties in the ranking of DMUs, three meta-frontier analyses are proposed; namely cross efficiency, super efficiency and lambda frequency. The applicability of the model is demonstrated on yearly *M3*–Competition time series data set. The proposed model can be implemented in forecasting software packages to help forecasters make a decision in regards to the best forecasting methods with conflicting results. One possible direction for future works is to develop an approach based on DEA to measure forecast accuracy and consequently select the best forecasting method across a cross-sectional and/or temporal hierarchy.

References

- [1] Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, **22** (4), 679-688.
- [2] Kitchenham, B. A., Pickard, L. M., MacDonell, S. G., & Shepperd, M. J. (2001). What accuracy statistics really measure [software estimation]. In *Software, IEE Proceedings-* (Vol. **148**, No. 3, pp. 81-85). IET.
- [3] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, S., & Winkler, R. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, **1** (2), 111-153.
- [4] Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, **16** (4), 451-476.
- [5] Syntetos, A. A., & Boylan, J. E. (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting*, **21** (2), 303-314.
- [6] Chatfield, C. (2013). *The analysis of time series: an introduction*. CRC press.
- [7] Yokuma, J. T., & Armstrong, J. S. (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods. *International Journal of Forecasting*, **11** (4), 591-597.
- [8] Athanasopoulos, G., & Hyndman, R. J. (2011). The value of feedback in forecasting competitions. *International Journal of Forecasting*, **27** (3), 845-849.
- [9] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2** (6), 429-444.
- [10] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, **30** (9), 1078-1092.
- [11] Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, **142** (1), 16-20.
- [12] Emrouznejad, A., & De Witte, K. (2010). COOPER-framework: A unified process for non-parametric projects. *European Journal of Operational Research*, **207** (3), 1573-1586.
- [13] Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA) Thirty years on. *European Journal of Operational Research*, **192**(1), 1-17.

- [14] Hyndman, R. J., & Athanasopoulos, G. (2014). Forecasting: principles and practice. *OTexts*.
- [15] De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, **22**(3), 443-473.
- [16] Collopy, F., & Armstrong, J. S. (1992). Expert opinions about extrapolation and the mystery of the overlooked discontinuities. *International Journal of Forecasting*, **8** (4), 575-582.
- [17] Witt, S. F., & Witt, C. A. (1992). *Modeling and forecasting demand in tourism*. Academic Press Ltd..
- [18] Geurts, M. D., & Patrick Kelly, J. (1986). Forecasting retail sales using alternative models. *International Journal of Forecasting*, **2** (3), 261-272.
- [19] Abdel-Khalik, A. R., & El-Sheshai, K. M. (1983). Sales revenues: Time-series properties and predictions. *Journal of Forecasting*, **2** (4), 351-362.
- [20] Fildes, R., Wei, Y., & Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, **27** (3), 902-922.
- [21] Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, **29** (3), 510-522.
- [22] Yokuma, J. T., & Armstrong, J. S. (1995). Beyond accuracy: Comparison of criteria used to select forecasting methods. *International Journal of Forecasting*, **11** (4), 591-597.
- [23] Brazdil, P. B., Soares, C., & Da Costa, J. P. (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning*, **50** (3), 251-277.
- [24] Nakhaeizadeh, G., & Schnabl, A. (1997). Development of Multi-Criteria Metrics for Evaluation of Data Mining Algorithms. In KDD (pp. 37-42).
- [25] Nakhaeizadeh, G., & Schnabl, A. (1998). Towards the Personalization of Algorithms Evaluation in Data Mining. In KDD (pp. 289-293).
- [26] Lim, D.-J., Anderson, T.R., Inman, O.L. (2014). Choosing effective dates from multiple optima in Technology Forecasting using Data Envelopment Analysis (TFDEA). *Technological Forecasting and Social Change* **88**, 91-97.
- [27] Inman, O.L., Anderson, T.R., Harmon, R.R. (2006). Predicting US jet fighter aircraft introductions from 1944 to 1982: A dogfight between regression and TFDEA. *Technological Forecasting and Social Change* **73**, 1178-1187.

- [28] Emrouznejad, A., Cabanda, E., & Gholami, R. (2010). An alternative measure of the ICT-Opportunity Index. *Information & Management*, **47** (4), 246-254.
- [29] Duong, Q. P. (1988). Model selection and ranking: An AHP approach to forecasts combination. *Mathematical and Computer Modelling*, **11**, 282-285.
- [30] Pendharkar, P. C., & Rodger, J. A. (2003). Technical efficiency-based selection of learning cases to improve forecasting accuracy of neural networks under monotonicity assumption. *Decision Support Systems*, **36** (1), 117-136.
- [31] Xu, B., & Ouenniche, J. (2011). A multidimensional framework for performance evaluation of forecasting models: context-dependent DEA. *Applied Financial Economics*, **21** (24), 1873-1890.
- [32] Hollingsworth, B., & Smith, P. (2003). Use of ratios in data envelopment analysis. *Applied Economics Letters*, **10** (11), 733-735.
- [33] Emrouznejad, A., & Amin, G. R. (2009). DEA models for ratio data: Convexity consideration. *Applied Mathematical Modelling*, **33** (1), 486-498.
- [34] Charnes, A., Cooper, W. W., Seiford, L.M., & Stutz, J. (1982) A multiplicative model for efficiency analysis. *Socio-Economic Planning Sciences*. *16*, 213, 24.
- [35] Banker, R. D., & Maindiratta, A. (1986). *Piecewise loglinear estimation of efficient production surfaces*. *Management Science*, **32** (1), 126-135.
- [36] Cook, W. D., & Zhu, J. (2013). DEA Cobb-Douglas frontier and cross-efficiency. *Journal of the Operational Research Society*, **65**(2), 265-268.
- [37] Chen, Y. (2005). Measuring super-efficiency in DEA in the presence of infeasibility. *European Journal of Operational Research*, **161** (2), 545-551.
- [38] Cook, W. D., Liang, L., Zha, Y., & Zhu, J. (2009). A modified super-efficiency DEA model for infeasibility. *Journal of the Operational Research Society*, **60** (2), 276-281.