COMPARISON OF AIR TRAVEL DEMAND FORECASTING METHODS

Ružica Škurla Babić, M.Sc.
Ivan Grgurević, B.Eng.
University of Zagreb
Faculty of Transport and Traffic Sciences
Vukelićeva 4, HR-10000 Zagreb, Croatia
skurla@fpz.hr, ivan.grgurevic@fpz.hr

Zvonimir Majić, B.Eng.
Pliva Croatia Ltd.
Prilaz baruna Filipovića 25, HR-10000 Zagreb, Croatia
zvonimir.majic@pliva.com

ABSTRACT

Accurate forecasts of future passenger demand are essential to effective revenue management system. The seat inventory control leans on predictions about the bookings to come to optimally allocate aircraft seats among the various booking classes. Forecasting for airline revenue management systems is inherently difficult because of complex nature of air travel demand which is highly stochastic. The problem is further complicated because of usually great number of origin destination pairs, each with its own seasonal and weekly effects, the economic environment and external factors like competition or special events. The paper describes general problem of forecasting airline demand and compares traditional methods of forecasting (moving averages, exponential smoothing, etc.) against neural networks as a forecasting method. All the methods are compared on the basis of standard error measures.

1 INTRODUCTION

Airlines forecast air travel demand in order to harmonize the complex set of their activities that will adequately match supply on the air transport market. For some business functions, the decisions are made based on the long-term traffic forecasts and then we speak about strategic decision-making. This, first of all, includes fleet planning, planning and evaluation of the flight networks and investment activities.

Tactical and operative decision-making is assisted by the mid- and short-term forecasts that are developed for the periods of six months to somewhat more than one year, leaning on different forecasting methodologies and different levels of demand aggregation.

Precise forecasting of the air traffic demand is especially important for efficient functioning of the airline revenue management systems that control the availability of travel seats in different booking classes with the goal of maximizing expected revenues. The demand forecasting module within the airline revenue management system generates the input data for the optimization module, i.e. data on the expected demand at the level of the price class of each flight. The estimates in practice suggest that the reduction of forecasting error by twenty percent results in the increase of the overall revenue on flight by 1 percent [1]. The airline revenue management mainly relies on the historical booking data of similar flights in order to estimate future demand.

The majority of quantitative methods that are described further in the text can be described as standard, i.e. methods that are used for forecasting in general. On the other hand, pick-up forecasting methods are used exclusively in airline revenue management systems. Instead of averaging the historical booking data, they calculate the increase, i.e. bookings
increment in time intervals between review points of the booking process. The increase for
the future periods is added to the number of on-hand bookings in order to estimate the number
of bookings at a certain moment in the future. In this context one should also mention the
forecasting using the method of linear regression which brings into connection the number of
confirmed bookings for a certain flight in the time intervals that precede the flight date and
the final number of bookings for that flight.

2 TIME SERIES FORECASTING METHODS

Historical data about different phenomena and in different research areas are usually
collected and analyzed in the form of time series with the aim of describing the phenomenon
that is being monitored, explaining its variations and predicting its movement in the future.

The methods of time series decomposition, methods of moving averages and various
smoothing methods belong to the methods of time series analysis with which it is possible to
forecast the level of phenomenon whereas various auto-regression and the respective models
measure the level of statistical relation between the members of the series, and are applied for
the description of phenomena that do not contain systematic components.

The statistical analysis of the movement of the level of a certain phenomenon over time
starts from the classical analysis of the time series into components:

- \( T_t \), trend component expresses the basic long-term tendency of phenomenon
development in time;
- \( C_t \), cyclical component expresses periodical repeating of certain values every two
  and more years;
- \( S_t \), seasonal component expresses fluctuations around the trend that are repeated
  in the similar way in the period which equals one year or less;
- \( e_t \), random, irregular or residual component expresses non-systemic influences
  on the phenomenon development, and it remains after removal of the systemic
  components (trend, seasonal and cyclical) [2].

2.1 Defining the sample for method comparison

Further in the text, a sample of 96 values (Table 1) that represent the monthly recorded
demand (in thousands) in a single airline during a time period of eight years, will be used to
show the results of the simple time series forecasting methods on the analysis. Although the
demand forecasting module of airline revenue management uses data on the number of
booking requests at micro level (exact flight and booking class), still the sample defined in
such a way, is suitable for elaboration of several methods of analysis and demand forecasting
since it contains the trend and seasonal components. The results have been calculated by
using the software package Zaitun Time Series (V 0.2.1.).

<table>
<thead>
<tr>
<th>Table 1: Sample statistical characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>characteristic</td>
</tr>
<tr>
<td>number of set elements</td>
</tr>
<tr>
<td>minimal value</td>
</tr>
<tr>
<td>maximal value</td>
</tr>
<tr>
<td>range</td>
</tr>
<tr>
<td>mean value</td>
</tr>
<tr>
<td>median</td>
</tr>
<tr>
<td>first quartile</td>
</tr>
<tr>
<td>third quartile</td>
</tr>
<tr>
<td>standard deviation</td>
</tr>
</tbody>
</table>
The methods will be compared by the most commonly used forecasting errors:

**Mean Absolute Error - MAE**

\[ MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t|, \tag{1} \]

**Mean Square Error - MSE**

\[ MSE = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2, \tag{2} \]

where \( T \) is the number of data that is used in the estimate, \( y_t \) observed value of the series at moment \( t \), and \( \hat{y}_t \) forecasted value of the series at moment \( t \).

### 2.2 Decomposition methods

Methods of time series decomposition base the forecasting on the separation of the basic components from the time series. For each component the forecasting of the future values is performed by forward extrapolation, and then by combining the separate forecasts the overall forecast is obtained.

In applying the additive model it is assumed that the seasonal and irregular component are independent of the trend, that the amplitude of seasonal variations does not change over time and that the annual average of seasonal fluctuations equals zero. The general form of additive model is:

\[ Z_t = T_t + C_t + S_t + e_t. \tag{3} \]

The multiplicative model of time series decomposition relies on the assumptions that the seasonal component amplitude is directly proportional to the trend level and that the irregular component variance is directly proportional to the value of systemic components. The general model of the multiplicative model multiplies the trend component with the coefficients of seasonal, cyclic and residual coefficient and has the form:

\[ Z_t = T_t \cdot I_{C_t} \cdot I_{S_t} \cdot I_{e_t}. \tag{4} \]

For the sample defined in the previous section, the analysis of the time series elements adaptation expectedly indicates the presence of the linear trend whose equation is:

\[ Z_t = 2.3194 \cdot x + 74.791, \tag{5} \]

and \( R \)-square value\(^1\) is 0.820617.

Graph 1 shows the linear trend and the forecasted demand values resulting from the introduction of the values for \( x \) for the next 12 months in the trend equation.

---

\(^1\) \( R \)-square value or determination coefficient is a number between 0 and 1 which indicates how well the estimated linear trend values correspond to actual data. The linear trend is most reliable when its \( R \)-square value is exactly or near 1.
The first step in the time series decomposition is the removal of the trend component. The forecast values resulting from the decomposition models can be seen in Graph 2 and the calculated seasonality coefficients are presented in Table 2.

Table 2: Seasonality coefficients of multiplicative and additive model of decomposition

<table>
<thead>
<tr>
<th></th>
<th>Iₜ₁</th>
<th>Iₜ₂</th>
<th>Iₜ₃</th>
<th>Iₜ₄</th>
<th>Iₜ₅</th>
<th>Iₜ₆</th>
<th>Iₜ₇</th>
<th>Iₜ₈</th>
<th>Iₜ₉</th>
<th>Iₜ₁₀</th>
<th>Iₜ₁₁</th>
<th>Iₜ₁₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>0,905</td>
<td>0,876</td>
<td>1,025</td>
<td>0,970</td>
<td>0,984</td>
<td>1,128</td>
<td>1,235</td>
<td>1,241</td>
<td>1,053</td>
<td>0,907</td>
<td>0,788</td>
<td>0,888</td>
</tr>
<tr>
<td>2</td>
<td>-17,70</td>
<td>-22,45</td>
<td>5,34</td>
<td>-4,78</td>
<td>-3,62</td>
<td>23,88</td>
<td>33,67</td>
<td>44,01</td>
<td>10,63</td>
<td>-14,20</td>
<td>-37,45</td>
<td>-17,33</td>
</tr>
</tbody>
</table>

2.3 Smoothing methods

The smoothing techniques are used for short-term forecasting, in the series with slight variations. Random or unpredictable influences of time series are smoothed and the last smoothed value is taken as the forecast value for the future periods.
2.3.1 Moving average methods

In statistics, the moving averages represent a series of data that have been calculated as simple or weighted averages of subsets of the basic set of data. The method of simple moving average is the simplest and easily applicable smoothing technique. The precise time series values for a certain period are substituted by the average of the respective value and several adjacent values ($M$ values).

The moving average method will react fast to major changes in the demand if $M$ is small. On the other hand, small $M$ results in estimates that are excessively sensitive to short-term random deviations of the values. In practice $M$ ranges between 2 and 15, and the very selection depends on the characteristics of the available data, length of the time intervals, and smoothing objective.

Graph 3 shows the moving averages for the defined time series sample. It is obvious that for a time series values with trend and seasonality, this method fails to be suitable.

![Graph 3: Forecasting by means of moving average method](image)

2.3.2 Exponential smoothing

The exponential smoothing methods belong to the mostly widespread demand forecasting methods in capacity management systems thanks to their simplicity, robustness, and precision. Simple exponential smoothing (SES) is the simplest exponential smoothing method, defined by the smoothing constant $\alpha$ which must be between 0 and 1. The forecast value $\hat{Z}_{t+1}$ for period $t + 1$ is calculated as the weighted average of the actual and forecast time series value in the previous time period $t$ in which the actual value $z_t$ is assigned the weight $\alpha$ and the forecast value $\hat{Z}_t$ is assigned the weight $1 - \alpha$, i.e.:

$$\hat{Z}_{t+1} = (1 - \alpha)\hat{Z}_t + \alpha z_t.$$  \hfill (6)

The $k$-period ahead forecast is given by:

$$\hat{Z}_{t+k} = \hat{Z}_{t+1} \quad k = 1, \ldots, K.$$  \hfill (7)

The recursive formula (6) can be written in the following way:

$$\hat{Z}_{t+1} = \alpha \sum_{j=0}^{\infty} (1 - \alpha)^j z_{t-j}.$$  \hfill (8)
Graph 4 shows the forecast values of the previously defined time series by the Simple exponential smoothing method. Using the software package Zaitun Time Series (V 0.2.1.), and the least square method, it was calculated that the least forecast error is given by the value $\alpha = 0.9$.

**Linear Exponential Smoothing (LES),** known as Holt's Method is used to smooth data that contain the linear trend. If we use $0 < \alpha < 1$ and $0 < \beta < 1$ to denote the smoothing parameters for $L$ and $T$, then the forecast for interval $t+1$ is given by the following formulas:

$$
\hat{Z}_{t+1} = L_t + T_t,
$$

$$
L_t = \alpha Z_t + (1-\alpha)(L_{t-1} + T_{t-1}),
$$

$$
T_t = \beta (L_t - L_{t-1}) + (1-\beta)T_{t-1}.
$$

While in simple exponential smoothing the forecast value is simply equal to the last value of $L$, in this case the recursive expression is given by:

$$
\hat{Z}_t+k = L_t + kT_t, \quad k = 1, \ldots, K.
$$

Graph 5 shows the forecast values of the previously defined time series by the exponential smoothing method with linear trend (LES). The values for $\alpha$ and $\beta$ have been calculated with Zaitun Time Series (V 0.2.1.) software, using the least square method.
Exponential smoothing method with trend and seasonality (Holt-Winter’s method-HW) is applicable in case when a series of data apart from trend contain also the seasonal component.

Let $0 < \alpha < 1$, $0 < \beta < 1$ and $0 < \gamma < 1$ be smoothing parameters for $L$, $T$ and $S$. Furthermore, we denote with $L$ the duration of the season in months, e.g. in case of monthly variations $L=12$, in case of half-a-year variations $L=6$.

Depending on the characteristics of the time series, the method is available in two versions: multiplicative and additive.

In multiplicative version, the forecast for interval $t + k$ has been set by the following expression [3]:

$$
\hat{Z}_{t+k} = (L_t + kT_t)S_{t+L-1}, \quad k = 1, \ldots, K,
$$

where the three components of forecast values are:

$$
\begin{align*}
L_t &= \alpha \left( \frac{z_t}{S_t} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}) \\
T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \\
S_t &= \gamma \left( \frac{z_t}{L_t} \right) + (1 - \gamma)S_{t-1}.
\end{align*}
$$

For the additive version the following expressions hold:

$$
\hat{Z}_{t+k} = A_t + kT_t + S_{t+L-1}, \quad k = 1, \ldots, K
$$

where the three forecast value components are [3]:

$$
\begin{align*}
A_t &= \alpha(z_t - S_{t-1}) + (1 - \alpha)(A_{t-1} + T_{t-1}) \\
T_t &= \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \\
S_t &= \gamma(z_t - A_t) + (1 - \gamma)S_{t-1}.
\end{align*}
$$

Graph 6 shows the forecast values of the previously defined time series using the Holt-Winter’s method, multiplicative and additive version. The values $\alpha$, $\beta$ and $\gamma$ have been selected by the least square method using software package Zaitun Time Series (V 0.2.1.).
For the calculation of the initial values $L_0$, $T_0$ and $S_0$ it is necessary to have available values $z_1$, $z_2$, …, $z_L$, that is, data for at least one year, and for the initial trend value one should know also the values of $z_{L+1}$ to $z_{2L}$, i.e. data for the second year. Initial values can be calculated in the following way [3]:

- for $L_0$ the average of the first year is taken:

$$L_0 = \frac{1}{L} \sum_{i=1}^{L} z_i \quad (15)$$

- for $T_0$ the average of the difference in the averages of the first and second year is taken:

$$T_0 = \frac{1}{L} \left( \frac{1}{L} \sum_{i=1}^{L} z_i - \frac{1}{L} \sum_{i=L+1}^{2L} z_i \right) \quad (16)$$

- the seasonality factor is calculated for $k=1, 2, \ldots, L$
  - for multiplicative version:

$$S_0(k) = \frac{1}{L_0} \left( z_k - (k-1) \cdot \frac{T_0}{2} \right) \quad (17)$$

  - for additive version:

$$S_0(k) = z_k - \left( L_0 + (k-1) \cdot \frac{T_0}{2} \right) \quad (18)$$

2.4 Neural network forecasting

The neural networks consist of two or more layers or groups of processing elements called neurons. The network processing capability is the consequence of the connections among these units, and it is achieved through the adaptation process or by learning from the set of learning examples. Neurons are connected into a network so that the output of every neuron is the input into one or several other neurons. The neurons are grouped into layers. Three basic types of layers are the input, hidden and output ones.

Standard error back-propagation algorithm includes optimisation of the error using the deterministic algorithm of the gradient descent. It calculates partial derivations of the quality criterion according to network parameters using recursive procedure which is performed reversely through the network from the output to the input network layer. The algorithm is based on the assumption that the error derivation propagation through the network is linear [4].

The results of applying the backpropagation multilayer feedfoward neural network on the previously defined sample using the software Zaitun Time Series are presented in Table 3 and Graph 7.

<table>
<thead>
<tr>
<th>Table 3: Summary of the applied neural network model</th>
</tr>
</thead>
<tbody>
<tr>
<td>value</td>
</tr>
<tr>
<td>Included observations</td>
</tr>
<tr>
<td>Input Layer Neurons</td>
</tr>
<tr>
<td>Hidden Layer Neurons</td>
</tr>
<tr>
<td>Output Layer Neurons</td>
</tr>
<tr>
<td>Output Layer Neurons</td>
</tr>
<tr>
<td>Learning Rate</td>
</tr>
<tr>
<td>Momentum</td>
</tr>
<tr>
<td>Iterations</td>
</tr>
</tbody>
</table>
2.5 Comparison of the accuracy of methods

Table 4 shows the forecasting error values MAE and MSE for the performed forecasting methods on the defined sample. It can be seen from the values of the forecasting errors for the method of multiplicative and additive time series decomposition that the multiplicative model of decomposition approximates more precisely the given time series than the additive model. The moving average method and the simple exponential smoothing method are not suitable for forecasting the time series values with trend and seasonality, and the best result for such a defined time series sample is obtained by the Holt-Winter’s exponential smoothing method, particularly the multiplicative version. The obtained results using the neural network are comparative to the multiplicative model of trend decomposition, and they are worse than the results obtained by Holt-Winter’s method.

![Graph 7: Examples of demand forecasting by neural network with “backpropagation” algorithm](image)

**Table 4: Forecasting errors of the model for a defined time series**

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>decomposition - multiplicative</td>
<td>10.16</td>
<td>158.11</td>
</tr>
<tr>
<td>decomposition - additive</td>
<td>12.45</td>
<td>266.59</td>
</tr>
<tr>
<td>Moving averages, M=2</td>
<td>22.64</td>
<td>802.67</td>
</tr>
<tr>
<td>Moving averages, M=5</td>
<td>28.62</td>
<td>1295.49</td>
</tr>
<tr>
<td>Moving averages, M=10</td>
<td>27.56</td>
<td>1389.14</td>
</tr>
<tr>
<td>Moving averages, M=15</td>
<td>27.48</td>
<td>12.91</td>
</tr>
<tr>
<td>SES, $\alpha = 0.9$</td>
<td>18.74</td>
<td>561.04</td>
</tr>
<tr>
<td>SES, $\alpha = 0.5$</td>
<td>21.46</td>
<td>761.54</td>
</tr>
<tr>
<td>SES, $\alpha = 0.1$</td>
<td>26.79</td>
<td>1355.37</td>
</tr>
<tr>
<td>LES</td>
<td>19.50</td>
<td>609.59</td>
</tr>
<tr>
<td>HW multiplicative version</td>
<td>7.65</td>
<td>97.56</td>
</tr>
<tr>
<td>HW additive version</td>
<td>9.44</td>
<td>140.13</td>
</tr>
<tr>
<td>neural network</td>
<td>10.14</td>
<td>179.57</td>
</tr>
</tbody>
</table>
3 APPLICABILITY OF ARTIFICIAL NEURAL NETWORKS FOR DEMAND FORECASTING IN AIRLINE REVENUE MANAGEMENT SYSTEMS

The professional literature provides multiple presentations of using the neural networks for time series forecasting with the results that justify further research and development of new algorithms. A minor number of studies has dealt with forecasting by means of neural networks on time series with pronounced seasonality, which are relevant within the context of the airline revenue management systems. The results of these studies lack uniformity. Whereas some authors advocate the application of neural networks on the data without prior de-seasonalisation, the others advocate precisely the opposite [5].

The characteristics of neural networks that contribute to their slow implementation in general, including then the limited use of neural networks in the forecasting models within the aircraft seat inventory control system are:

- neural networks are computationally very demanding, output of every neuron being the result of adding several products and calculating the non-linear activation function;
- neural networks require large memory space since each neuron has several synaptic connections, whose weight coefficient has to be stored in the memory; the neural network memory requirements grow with the square of the neuron number;
- the computational speed of the neural network is determined by the number of mathematical operations of a single neuron, rather than the complete network since every network layer has parallel structure, and every neuron in a layer may be observed as a local processor that works parallel with other neurons [6].

Parallel to the studies of the structures of neural networks and the models of synaptic connections, as well as the development of the learning algorithm, the methods of their implementation that ensure optimal usage of the neural network properties have also been studied. The majority of applied neural networks has been implemented on conventional computer systems that had not been designed exclusively for the implementation of neural networks, for which more adequate solutions are those that use parallel structure of the neural networks.

The real usage of all the good properties of the neural networks can be expected only when good hardware is available, specialized for their implementation, so that the core of the research activities in this area is focused on the development of specialized electronic and optical, i.e. optoelectronic implementations. Electronic implementations of neural networks are based on the bus-oriented processes, coprocessors, CCDs (Charge Coupled Device technology) and VLSI (Very Large Scale Integrated) assemblies, and optical/optoelectronic implementations on optical or combined optical and electronic components [6].

In their paper which deals precisely with forecasting of transport demand based on the data about the realized transport of one airline – and which is characterised by the so-called multiplicative seasonality, J. Faraway and C. Chatfield emphasise as especially important the selection of:

- adequate set of input variables and weights;
- adequate network architecture;
- adequate activation functions that are not to be equal in the hidden and the output layer;
- adequate numerical procedure for the neural network model calibration [7].

Zhang and Kline carried out a comprehensive research and analyzed 48 models of neural networks on a large set of data (756) with seasonal variations and concluded that as a
rule the simple models surpass the complex models and that the efficiency is most improved by previous “cleaning” of data from the trend and season components [8].

This paper has presented the application of neural networks on time series data with trend and, seasonality, and the obtained results indicate good forecast accuracy.

4 CONCLUSIONS

The dominant demand forecast values for air travel demand are simple time series models such as the moving average or the exponential average smoothing, with which the known values, i.e. historical data are averaged and adapted to the known seasonal impacts. The sophisticated time series models, such as e.g. autoregressive moving average isolate among the historical data the trends and forms of demand fluctuation and extrapolate from them the market trend forecasts. More complex models, such as e.g. multiple regression, Kalman’s filters and neural networks are seldom used since the specific dynamic of every OD pair and the necessity for the definition of the exact interaction between the variables on every market makes the construction and upgrading of such models a long-term one.

The neural networks, although being very sophisticated tools for time series forecasting with seasonality, leave a lot of room for errors as e.g. non-convergence, convergence into local minimum or may even result in unreasonable forecasts. By adding hidden layers into the neural network model, the number of network parameters is increased, which may yet lead to wrong forecasting. Apart from the previously mentioned reasons, the modest share of neural networks in airline revenue management systems is not surprising.

LITERATURE

1. Pölt, S.: Forecasting is difficult – especially if it refers to the future, AGIFORS Reservations and Yield Management Study Group Annual Meeting Proceedings, Melbourne, Australia, 1998
6. Petrović, I., Perić, N.: Inteligentno upravljanje sustavima, University of Zagreb, Faculty of Electrical Engineering and Computation, Zagreb, 2008, p. 15