ESTIMATION OF AUSTRALIA’S OUTBOUND AIRLINE PASSENGER DEMAND USING AN ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

Panarat Srisaeng¹, Glenn Baxter²
¹,² School of Tourism and Hospitality Management, Suan Dusit University, Hua Hin Campus 77110, Prachaup Khiri Khan, Thailand

Received 9 January 2021; accepted 22 March 2021

Abstract: This study has proposed and empirically tested an adaptive neuro-fuzzy inference system (ANFIS) model for predicting Australia’s outbound international airline passenger demand. The model was developed using eleven input parameters of world GDP, world population, world air fare yields, world jet fuel prices, outbound flights from Australia, Australia’s unemployment numbers, Australian’s (AUD/USD) foreign exchange rate, Australia’s outbound tourist expenditure and four dummy variables. The model was constructed using annual data from 1994 to 2019. The hybrid learning algorithm and the subtractive clustering partition method were used to generate the optimum ANFIS models. The performance of the model was measured using five error measures: coefficient of determination (R²-value), root mean square errors (RMSE), mean absolute errors (MAE) and the mean absolute percentage error (MAPE). The results found that the mean absolute percentage error (MAPE) for the overall data set of the model was 3.60%. The R²-value was around 0.9886, demonstrating that the ANFIS is an efficient model for predicting Australia’s outbound airline passenger demand.

Keywords: Australia, adaptive neuro-fuzzy inference system, airlines, ANFIS, forecasting, hybrid learning algorithm, outbound passengers, subtractive clustering technique.

1. Introduction

Due to the Australia’s size and relatively remote geographical location, the airline industry plays a vital role in linking the country with key air travel markets all around the world. These air transport services help facilitate commerce, trade, and tourism. In 2019, sixty-four international scheduled airlines, including 5 dedicated all-cargo airlines, operated services to/from Australia. A total of 21.1 million passengers travelled from Australia by air in 2019 (Bureau of Infrastructure, Transport and Regional Economics, 2020). Like other international air travel markets, Australia’s outbound air services are provided by both full-service network carriers (FSNCs), for example, Qantas, Cathay Pacific Airways and Thai International, as well as by low-cost carriers (LCCs), such as AirAsia-X, Cebu Pacific Air, Jetstar Airways, and Scoot.

Both full-service network carriers (FSNCs) and low-cost carriers (LCCs) require highly accurate passenger forecasts as these are...
used in their corporate and annual operating plans, fleet-planning and aircraft acquisition, flight scheduling and their route network development (BaFail et al., 2000; Doganis, 2019). Future passenger traffic forecasts are also used by airlines to develop their marketing strategies and promotional programs and to establish the public’s future requirements, desires, and ability to travel (Radnoti, 2002). Future passenger forecasts also play a key role in an airline’s risk management function through an objective evaluation of the demand side of their business (Abed et al., 2001; BaFail et al., 2000). In the airline industry virtually every tactical or strategic decision made by an airline ultimately comes from a forecast (Doganis, 2019).

The aim of this study is to apply an adaptive neuro-fuzzy inference system (ANFIS) for the estimation of Australia’s outbound airline passenger demand. The proposed model could be used by either the FSNCs or LCCs serving Australia’s international air travel market. In this paper, a fuzzy rule-based model was analyzed for estimating Australia’s outbound international airline passenger demand.

The remainder of the paper is structured as follows: Section 2 presents a literature review of the relevant studies, focusing on the approaches used to predict air passenger demand. Section 3 outlines the methodology underpinning the study. The proposed ANFIS method and a real-world case study focusing on forecasting Australia’s outbound air passenger demand is presented in Section 4. Finally, the study’s findings are presented in Section 5.

2. Materials and Methods

2.1. Dataset and Variables Selection

The study period was from 1994 to 2019. There were 8 variables included in the data set. The first variable was world real GDP, which was included in the model as a measure of the monetary value of the world economy (Cook and Billig, 2017; Janić, 2007). The world real GDP data was sourced from the International Monetary Fund. The second variable included in the modelling was world population, which measured the size of the world population. Population size is a determinant of air travel demand (BaFail et al., 2000). The world population data was also sourced from the International Monetary Fund. World jet fuel prices was the third variable included in the model as these have been shown to influence air travel demand (Klophaus, 2009). The world jet fuel prices data was obtained from the United States Energy Information Agency. The price of air trips was included in the modelling using world air fare yields data from Boeing Commercial Airplanes (Profillidis and Botzoris, 2019). Australia’s unemployment, as measured by the number of unemployed persons, was included in the modelling. Unemployment rates have been acknowledged as a determinant of air travel demand (Clark et al., 2009; Wensveen, 2015). Australia’s unemployment data was sourced from the Australian Bureau of Statistics. Australia’s outbound tourist expenditure was sourced from the World Bank as tourism expenditure has an impact on air travel demand (Profillidis and Botzoris, 2019).
Australia’s annual exchange rate with the United States dollar (USD) was included in the modelling. Foreign exchange rate is a valid factor that influences international air travel demand (BaFail et al., 2000; Dargay and Hanley, 2001). Australia’s annual Australian dollar/United States dollar exchange rate data was obtained from the Reserve Bank of Australia. There is an extraordinarily strong relationship between tourism and the demand for air transport services (Duval, 2019; Stabler et al., 2010). Considering this important relationship, tourism expenditure was included in the modelling to account for the influence that tourism attractiveness has on airline passenger demand (Srisaeng et al., 2015a, 2015b). Australia’s tourism data was sourced from the Australian Bureau of Statistics. The final independent variable included in the modelling was Australia’s outbound flights, which measures the annual number of flights operated from Australia (Garrow, 2010). Australia’s outbound flights data was sourced from the Bureau of Infrastructure, Transport and Regional Economics (BITRE). The output variable was Australia’s annual enplaned international passengers. This data was also sourced from the Bureau of Infrastructure, Transport and Regional Economics (BITRE).

It was necessary to include four dummy variables in the ANFIS Modelling: the Sydney Olympic Games held in 2000. The Olympic Games ran for 17 days (the Opening Ceremony was held on 15 September 2000). The Paralympics were also staged in Sydney over a 12-day period shortly after the conclusion of the Olympic Games (Madden, 2002). During the Sydney Olympic Games, extra international flights were operated to satisfy the additional passenger and tourism demand associated with the staging of the Olympic Games (Hensher and Brewer, 2002).

The second dummy variable controlled for the impact of 9/11 and the collapse of Ansett Australia in 2001. Passenger traffic in 2002 was impacted by the terrorist attacks on the USA in September 2001 (Jaffe, 2016; Janić and Stough, 2005). The collapse of Ansett Australia (also in September 2001) may have also influenced Australia’s outbound passenger numbers in 2002 (Prideaux, 2003).

The third dummy variable (Dummy 3) controlled for the influence of the Commonwealth Games, which were held in Melbourne from 15 to March 26, 2006. The 2006 Melbourne Commonwealth Games was the largest sporting and community event held in Victoria’s history (KPMG, 2006). International airlines played a key role in carrying competitors and visitors to Melbourne, so they could attend Commonwealth Games events.

The fourth dummy variable (Dummy 4) controlled for the influence of the second Commonwealth Games, which were held in Gold Coast, Queensland from 4 to April 15, 2018 (Commonwealth Games, 2020).
2.2. Data Normalization

Before commencing the data training in the adaptive neuro-fuzzy inference system (ANFIS), it is necessary to process the data into patterns. This practice ensures that the ANFIS will be trained effectively and will alleviate any variable skewing the results significantly (Srisaeng et al., 2015a, 2015b). Data normalization ensures that all the input parameters are of equal importance in training the artificial neural network (ANN) system within the ANFIS (Baseri, 2011).

In the data normalization process, the study’s data are scaled so they fall within a pre-specified range, such as [0, 1] (Mitsa, 2010; Narang et al., 2017). In the present study, all data were normalized prior to being used training phase using Eq. 1.

\[
x_{\text{norm}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \tag{1}
\]

2.3. The Architecture of the Adaptive Neuro-Fuzzy Inference System (ANFIS)

The architecture of an adaptive neuro-fuzzy inference system is comprised of the fuzzy inference system (FIS) and an artificial neural network with given input and output data pairs (Al-Mayyahi et al., 2014; Kaleel and Mallick, 2012; Shukla et al., 2010). The adaptive neuro-fuzzy inference system enables the fuzzy logic capability to adapt to the identified membership parameters that best enable the associated fuzzy inference system (FIS) to track the given input and output data parameters of the ANFIS model (Srisaeng et al., 2015a, 2015b). To present the ANFIS architecture, two fuzzy if-then rules based on a first order Takagi-Sugeno model are considered in the present study (Al-Mayyahi et al., 2014; Lerkkasemsan, 2017; Srisaeng et al., 2015a):

Rule 1: if \( x \) is \( A1 \) and \( y \) is \( B1 \) then \( f1=p1 \, x+q1 \, y+r1 \);
Rule 2: if \( x \) is \( A2 \) and \( y \) is \( B2 \) then \( f2=p2 \, x+q2 \, y+r2 \).

Here, \( x \) and \( y \) are the inputs, \( Ai \) and \( Bi \) are the fuzzy sets, \( fi \) are the outputs within a fuzzy rule, and \( p_i, q_i, r_i \) are the design parameters which are determined during the training process (Zendehboudi et al., 2017). Figure 1 depicts the ANFIS model architecture.

![The Adaptive Neuro-fuzzy Inference System Architecture](image)

Fig. 1. The Adaptive Neuro-fuzzy Inference System Architecture
The ANFIS is comprised of five layers: Layer 1 is the fuzzification layer, in which the names of the fuzzy sets are defined by the following equation:

\[ O^1_i = \mu A_i(x), i = 1,2 \text{ or } O^1_i = \mu B_{i-2}(y), i = 3,4 \quad (2) \]

Where \( x \) and \( y \) are the input to the \( i \)th node and \( A_i \) and \( B_{i-2} \) are linguistic labels associated with this fuzzification node (Übeyli et al., 2010).

Layer 2 is a rule layer. In this layer each node multiplies the incoming inputs and subsequently sends the results as an output. The result is the firing strength of the node (Xiao et al., 2014). Also, in Layer 2 the weight functions \( w_i \) for the following layer is defined (Vahabi et al., 2016).

\[ O^2_i = w_i = \mu A_i(x) \times \mu B_i(y), i = 1,2 \quad (3) \]

Layer 3 is the normalization layer. In this normalization layer the ratio of the \( i \)th rules firing strength is calculated from the sum all of all the rules firing strengths (Xiao et al., 2014).

\[ O^3_i = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1,2 \quad (4) \]

Layer 4 is the defuzzification layer in which the nodes are adaptive nodes with a node function (Srisaeng et al., 2015b; Zendehboudi et al., 2017).

\[ O^4_i = \bar{w}_i f_i = \bar{w}_i (p_1x + q_1y + r_1), i = 1,2 \quad (5) \]

The fifth ANFIS layer, whose node is labelled “\( \Sigma \)”, is the output layer, in which a single node calculates the overall output as a summation of all incoming signals (Srisaeng et al., 2015a, 2015b; Übeyli et al., 2010).

\[ O^5_i = \sum_i w_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (6) \]

In Layer 1, the parameters are referred to as the premise parameters (Savkovic et al., 2019) whilst the parameters in Layer 4 are called the consequent parameters in Layer 4 (Kumar and Vaidehi, 2017).

2.4. Proposed ANFIS for Predicting Australia’s International Outbound Passenger Air Travel Demand

This study used the Takagi-Sugeno fuzzy model in the ANFIS modelling of Australia’s outbound airline passenger demand. The Takagi-Sugeno fuzzy model uses a mixture of the back propagation to learn the membership functions and least mean square estimation to determine the coefficients of the linear combinations in the fuzzy rule conclusions. There are two phases involved with the learning process. In the first phase the input patterns are propagated, and the optimal conclusion parameters are subsequently estimated through an iterative least square procedure, whilst the antecedent parameters (membership functions) are assumed to be in a fixed state for the current cycle through the training data set. In the second phase the patterns are propagated again, and within this epoch, the back propagation is utilized to modify the membership functions, while the conclusion parameters remain fixed. This function is then iterated (Jiang et al., 2018, p.59).

There are several methods available to classify the input data and for the fuzzy rulemaking, among which the most common being the grid partition and subtractive fuzzy clustering. In this study
due to the number of input variables and the requirement for considerable membership functions, the subtractive clustering method was utilized. Hence, subtractive fuzzy clustering was used to establish the rule-based relationship between the input and output variables (Abraham, 2005; Srisaeng et al., 2015a).

The first step in the ANFIS modelling involved the generation of the fuzzy inference system (FIS) (Srisaeng et al., 2015a, 2015b). In the subsequent step, the FIS parameters from the training datasets were then optimized. This was achieved using the least square method and the back-propagation gradient descent method for training the ANFIS model (Wei et al., 2011; Zaki et al., 2012). The training of the study’s data was subsequently performed automatically in the ANFIS system. In this process a range of training errors was obtained for evaluation (Yetilmezsoy et al., 2011). Following the conclusion of the ANFIS training, an ANFIS model with forecasting function was obtained for the output forecasting. Finally, a performance index was developed to evaluate the performance of the models that were developed during the ANFIS training process (Srisaeng et al., 2015a, 2015b).

The Takagi Sugeno ANFIS network setup process was developed with 12 Gaussian membership type functions. The ANFIS model developed for the present study used the hybrid learning algorithm. The architecture of the study’s ANFIS is presented in Figure 2.

Fig. 2.
The Optimum ANFIS Model Architecture for Forecasting Australia’s Outbound Airline Passenger Demand

The adaptive neuro-fuzzy inference system models were run for each combination of model parameter with varying numbers of epochs to avoid the possible over-fitting of the model (Chen et al., 2010). The Gaussian-curve membership function and 15 rules are the optimum architecture for the Australia’s outbound airline passenger demand ANFIS model.

2.5. Training the ANFIS Model

In ANFIS modelling the training process is a vital part of the ANFIS model development process. The training process begins with the receipt of the training input/output data set. It is a requirement to use two vectors to train the ANFIS as the training data comprises a set
of input and output vectors. The training data set is used to ascertain the parameters for the membership function (MF) (Srisaeng et al., 2015a). Accordingly, the training process is used to optimize the ANFIS model (Suparta and Alhasa, 2016). The testing process checks the model’s performance, and hence, the generalization ability of the developed model (Kovač et al., 2020).

The testing data subset was kept separate to the training data set and was used to train the ANFIS model (Srisaeng et al., 2015a, 2015b). The data in this study was therefore split into two selected groups: the first group of 20 data was used as the training set (80% of the overall data), and the remaining 6 data was used for testing and validate the robustness of the ANFIS-based prediction model (Srisaeng et al., 2015a; Wei et al., 2011).

The study’s ANFIS model used 20 training data in 1-400 training epochs (Wei et al., 2011). Figure 3 shows the training curve of ANFIS model with root mean square error (RMSE) of 0.0000019.

Fig. 3.
Error Change during Training the ANFIS Outbound Passenger Model

The actual and ANFIS Australia’s outbound passenger predicted values following the completion of training are presented in Figures 4. Figure 4 shows that the ANFIS system is well-trained to model Australia’s outbound airline passenger demand.

Fig. 4.
Actual and Predicted Australia’s Outbound Passenger Values
2.6. Testing of the ANFIS Model

The testing of the ANFIS model was completed using the testing dataset which, as previously noted, was kept separate to the model (Mehta and Jain, 2009; Srisaeng et al., 2015a). The testing process ensured that the ANFIS model had been trained and was able to capture all the various characteristics of the targets, and hence, avoid the possible overfitting of the model (Chen et al., 2010).

2.7. ANFIS Model Goodness of Fit Measures

Four goodness of fit measures were used in the study to evaluate the performance of the ANFIS models, the Root Mean Squared Error (RMSE), mean absolute error (MAE), the mean absolute percentage error (MAPE), mean square error (MSE), and coefficient of determination (R²), were calculated using Eq. (7) – Eq. (11) (Srisaeng et al., 2015a, 2015b):

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (t_i - \hat{t}_i)^2}
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |t_i - \hat{t}_i|
\]

\[
MAPE = \frac{1}{N} \left( \sum_{i=1}^{N} \frac{|t_i - \hat{t}_i|}{t_i} \right) \times 100
\]

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (t_i - \hat{t}_i)^2
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (t_i - \hat{t}_i)^2}{\sum_{i=1}^{N} (t_i - \bar{t})^2}
\]

Where \( t_i \) is the actual values, \( \hat{t}_i \) is the predicted values, \( N \) is the total number of data (Tiryaki and Aydin, 2014, p. 104).

3. Results

The optimum ANFIS model architecture for forecasting of Australia’s outbound airline enplaned passengers is shown in Figure 5.

![Fig. 5. Australia's Outbound Airline Passenger ANFIS Forecasting System Structure](image)

The ANFIS was trained using Matlab R2020a using various possible combinations of the subtractive clustering parameters (range of influence (ROI) = 0.45-0.60, squash factor (SF) = 1.20-1.35, accept ratio (AR) = 0.40-0.55 and reject ratio (RR) = 0.10-0.20) for the range of epoch number from 1-400 epochs. The developed ANFIS model was manipulated until the best settings were obtained based on the
lowest RMSE value. As noted earlier, data normalization was undertaken to increase the training performance of the model. The training process was completed whenever the maximum epoch number was reached, or the training error goal was achieved (Srisaeng et al., 2015a, 2015b).

In this study, the root mean square errors (RMSE) became steady after running 2 epochs of training data. The final convergence values were 0.000002.

The ANFIS model was manipulated by changing the parameters of clustering systematically around their default values until the best settings were obtained based on the lowest RMSE value. It was found that the optimum ANFIS model with ROI = 0.50, SF = 1.25, AR = 0.50 and RR = 0.15 returns the lowest value of RMSE at 0.00000187.

Upon conclusion of the training phase, the ANFIS model for predicting Australia’s outbound passenger was tested and validated by selecting 6 data points, which are different from the other 20 points used for ANFIS training (Srisaeng et al., 2015a, 2015b). Each validation data point was fed into the ANFIS system and then Australia’s forecasted outbound passenger values were computed and compared to the actual values.

Figure 6 presents examples of the surface graphs obtained from the ANFIS. These graphs show the variation of output with respect to two various parameters (X and Y-axis).

Fig. 6.
Obtained Surfaces in ANFIS Model: Outbound Passengers vs. International Passenger Yields, World GDP, and World Jet Fuel Prices

The performance index of training, testing, validating and overall data of Australia’s outbound airline passenger model were calculated as shown in Table 1. Table 1 shows that Australia’s outbound passenger ANFIS model has a very satisfactory predictive capability. The model shows that MAE, MAPE, MSE, RMSE are very low for training, testing, validating and overall data sets.
Table 1
The Training, Testing, and Overall Data Set Performance Indexes of the Australian Outbound Airline Passenger Demand ANFIS Model

<table>
<thead>
<tr>
<th>Performance Index</th>
<th>Training Data</th>
<th>Testing Data</th>
<th>Validating Data</th>
<th>Overall Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.0020</td>
<td>0.0242</td>
<td>0.1116</td>
<td>0.0171</td>
</tr>
<tr>
<td>MAPE</td>
<td>1.06%</td>
<td>11.73%</td>
<td>12.02%</td>
<td>3.60%</td>
</tr>
<tr>
<td>MSE</td>
<td>$1 \times 10^{-5}$</td>
<td>0.0009</td>
<td>0.013</td>
<td>0.0016</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.002</td>
<td>0.030</td>
<td>0.1155</td>
<td>0.041</td>
</tr>
</tbody>
</table>

The overall estimated and actual value of Australia’s outbound passenger was regressed and as Figure 7 shows the coefficient of determination ($R^2$) is high, being around 0.9886.

Fig. 7.
Comparison of Estimated and Actual Values of the ANFIS Model for Forecasting Australia’s Outbound Passenger Demand

The actual and forecasted values of Australia’s outbound passenger demand model are plotted in Figure 8. Figure 8 clearly highlights the very good fit of the ANFIS to the actual data, indicating the high estimation accuracy of the study’s ANFIS model.

Fig. 8.
A Comparison of Australia’s Actual and Forecasted Outbound Airline Passengers: 1994-2019
4. Conclusions

One of the most critical areas of airline management is the prediction of future air travel demand as these forecasts help guide a wide range of very important management decisions and plans. The primary objective of this study was for the first time to propose and empirically test an adaptive neuro-fuzzy inference system (ANFIS) model for estimating Australia’s outbound international airline passenger demand. The study covered the period 1994 to 2019. The key input variables in the ANFIS modelling were world GDP, world population, world jet fuel prices, world air fares (proxy for air travel cost), Australia’s tourism attractiveness, outbound flights, Australia’s unemployment levels, the Australian and United States foreign exchange rate and four dummy variables that controlled for major exogenous events that had impacted Australia’s outbound airline passenger demand over the study period. The dummy variables were incorporated to control for the influence of the Sydney Olympic Games in 2000, the tragic events of 9/11, the Commonwealth Games, which were held in Melbourne in 2006, and the Commonwealth Games held in the Gold Coast during 2018.

Sugeno fuzzy rules were used in the ANFIS structure and Gaussian membership function and linear membership functions were also developed. The hybrid learning algorithm and the subtractive clustering partition method were used to generate the optimum ANFIS models. Data was normalized to the scale [0,1] to increase the model’s training performance. The results found that the mean absolute percentage error (MAPE) for the overall data set of Australia’s outbound passenger demand model was 3.60%. The study concludes that the proposed ANFIS model is very promising for predicting air travel demand.

References


