FORECASTING DEMAND FOR LOW COST CARRIERS IN AUSTRALIA
USING AN ARTIFICIAL NEURAL NETWORK APPROACH

Panarat SRISAENG, Glenn S. BAXTER, Graham WILD

School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University,
Melbourne, Australia 3001
E-mails: s3125221@student.rmit.edu.au1 (corresponding author); glenn.baxter@rmit.edu.au2; graham.wild@rmit.edu.au3

Received 21 July 2014; accepted 23 March 2015

Panarat SRISAENG, PhD (Candidate)
Education: Bachelor of Economics, Chulalongkorn University, Bangkok, Thailand, 1993.
Master of Business Economics, Kasetsart University, Bangkok, Thailand, 1998.
Affiliations and functions: PhD (candidate) in aviation, RMIT University, School of Aerospace, Mechanical and Manufacturing Engineering and a lecturer in airline business management at Suan Dusit Rajabhat University, School of Tourism and Hospitality Management, Thailand.
Research interest: low cost airlines management; demand model for air transportation; demand forecasting for air transportation.

Glenn S. BAXTER, PhD (Aviation)
Education: Bachelor of Aviation Studies, University of Western Sydney, Australia, 2000.
Master of Aviation Studies, University of Western Sydney, Australia, 2002. PhD, School of Aviation, Griffith University, Brisbane, Australia, 2010.
Affiliations and functions: lecturer in Aviation Management and deputy director Undergraduate Aviation Programs, RMIT University, School of Aerospace, Mechanical and Manufacturing Engineering.
Research interest: air cargo handling and operations, airport operations, sustainable aviation, supply chain management.
Publications: author of 7 scientific articles, 5 conference papers.

Graham WILD, PhD Eng.
Date of birth: 1981.
Research interest: aircraft systems, aerospace vehicle health monitoring, optical fibres, sensing and instrumentation, sustainable aviation, aviation technology.
Publications: author of 15 articles, 4 book chapters, 60+ conference presentations.
Present position: 2013 – Aviation Program Manager, RMIT University, School of Aerospace Mechanical and Manufacturing Engineering.
Abstract. This study focuses on predicting Australia’s low cost carrier passenger demand and revenue passenger kilometres performed (RPKs) using traditional econometric and artificial neural network (ANN) modelling methods. For model development, Australia’s real GDP, real GDP per capita, air fares, Australia’s population and unemployment, tourism (bed spaces) and 4 dummy variables, utilizing quarterly data obtained between 2002 and 2012, were selected as model parameters. The neural network used multi-layer perceptron (MLP) architecture that compromised a multi-layer feed-forward network and the sigmoid and linear functions were used as activation functions with the feed forward-back propagation algorithm. The ANN was applied during training, testing and validation and had 11 inputs, 9 neurons in the hidden layers and 1 neuron in the output layer. When comparing the predictive accuracy of the two techniques, the ANNs provided the best prediction and showed that the performance of the ANN model was better than that of the traditional multiple linear regression (MLR) approach. The highest $R$-value for the enplaned passengers ANN was around 0.996 and for the RPKs ANN was round 0.998, respectively.

Keywords: air transport, artificial neural network (ann), Australia, forecasting methods, low-cost carrier.

1. Introduction

The emergence of low cost carriers (hereafter LCCs) has become a global phenomenon, with today virtually all travel markets containing at least some low cost carriers (Vasigh et al. 2008). Following the deregulation of Australia’s domestic airline market in 1990, which permitted other airlines to compete with the established carriers (Forsyth 2003; Nolan 1996), a number of low cost carriers (LCCs) have entered the market. The LCCs now have around a 35 per cent market share, with the two major incumbent LCCs being Jetstar and Tiger Airways (Srissaeng et al. 2014).

Forecasting is considered the most critical area of airline management. Airlines forecast demand in order to plan the supply of services that are necessary to satisfy that demand (Dogans 2009). Forecasting passenger transport demand is, therefore, of critical importance for airlines as well as for investors, since investment efficiency is greatly influenced by the accuracy and adequacy of the estimation performed (Blinova 2007). Air traffic forecasts are one of the key inputs into an airline’s fleet planning, route network development, and are also used in the preparation of the airline’s annual operating plan (Ba-Fail et al. 2000; Doganis 2009). Furthermore, analysing and forecasting air travel demand may also assist an airline in reducing its risk through an objective evaluation of the demand side of the airline business (Ba-Fail et al. 2000). Hence, the sustainable success of any firm is closely related to the ability of its management or decision makers to foresee the future and define and implement appropriate strategies (Sivrikaya, Tunç 2013).

Consequently, the forecasting of LCC demand plays an important role in decision making and planning for both airlines and airports. In the past, regression models were generally used to predict air traffic demand (see, for example, Abed et al. 2001; Aderamo 2010; Ba-Fail et al. 2000; Bhadra 2003; Kopsch 2012; Sivrikaya, Tunç 2013). While the regression analysis method is supported by statistical theories as producing good estimates (according to certain statistical properties), for instance, being the best linear unbiased estimator, other approaches such as artificial neural network (ANN) have been found to be very useful in developing predictive models in other fields (Alekseev, Seixas 2009; Tso, Yau 2007).

In recent years, the use of artificial neural networks has grown rapidly due to its ability of mapping any linear or non-linear function and having no associated data assumption requirements (Claveria, Torra 2014; Kunt et al. 2011; Santos et al. 2014). This technique consists of training a computer to learn from substantial data based on the structure of human brain, using many simple processing elements (Haykin 1999). This approach has now been applied to a wide range of disciplines, including transportation (Jiménez et al. 2014), banking (Venkatesh et al. 2014), energy demand prediction (An et al. 2014; Jarimillo-Morán et al. 2013), tourism demand forecasting (Claveria, Torra 2014; Palmer et al. 2006), traffic accident prediction (Akgüngör, Doğan 2009; Kunt et al. 2011), and economics (Choudhary, Haider 2012).

The primary advantage of an ANN over other forecasting methods is that the network equally well predicts the processes whose regular components have any distribution law, whereas most other forecasting methods are best suited for processes that possess a regular component that belongs to a specific class (clearly, the method of polynomial smoothing is best suited for processes with a polynomial regular component, the method of smoothing by Fourier series is best suited for processes with a periodic regular component and so forth). A further advantage of ANNs is their ability to learn (Aizenberg 2011; Mrugalski 2013; Sineglazov et al. 2013).

This paper proposes and empirically tests for the first time classical linear regression and artificial neural network (ANN) models that can be used to forecast Australia’s low cost carrier (LCC) quarterly passenger demand, as measured by enplaned passengers and revenue passenger kilometres performed (RPKs).

2. Forecasting models for estimating air transport passenger demand

While the traditional regression forecasting method has its own model assumptions and pre-defined underlying relationships between dependent and independent variables (explanatory), such as normal data, linear
relationships between dependent and independent variables, low multicollinearity, which may often be violated (Garrido et al. 2014), neural network models are considered a superior forecasting method since there are no prior assumptions about the underlying patterns in the data in the model development process (Garrido et al. 2014; Pan et al. 2013). Furthermore, the major conceptual limitation of all regression techniques is the fact that the researcher is only able to ascertain relationships, but may not be sure about underlying causal mechanisms in the model (Aparicio, Aparicio-Ruiz 2002; Tso, Yau 2007).

In the air transport industry, many services providers and government regulatory agencies follow the International Civil Aviation Organization (ICAO) Manual on Air Traffic Forecasting. This manual was originally developed in 1985 using traditional modelling techniques (Alekseev, Seixas 2009). There have been few reported studies that have used neural networks in air transport demand forecasting. Attempts at using neural network modelling applications for estimating air transport demand have been reported by Alekseev and Seixas (2002, 2009), who developed neural network based forecasting models to predict the annual Brazilian air transport passenger demand. In a further study, Chen et al. (2012) employed a back-propagation neural network (BPN) to improve the forecasting accuracy of air passenger and air cargo demand from Japan to Taiwan.

2.1. Classical modelling approach

The standard approach used in air transport modelling and forecasting is to define it as a process in which two vectors from different domains (a vector built from socio-economic variables and a further vector built from air transport system variables) can be combined through mapping, so that the target demand forecasting can be accomplished (Alekseev, Seixas 2009; Rengaraju, Arasan 1992).

Multiple linear regression (MLR) is a linear statistical technique that is very useful for predicting the best relationship between a dependent variable and several independent variables. MLR is based on least squares: the model is fit so that the sum of squares of differences of observed and predicted values is minimized (Tiryaki, Aydın 2014).

A general MLR model can be formulated as per the following equation:

\[ Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_i X_i + \varepsilon, \]

where \( Y \) indicates the dependent variable, \( X_i \) represents independent variables, \( \beta_i \) represents predicted coefficients, and \( \varepsilon \) is the error term (Tiryaki, Aydın 2014).

The basic econometric model tested in this study uses six inputs: the GDP (Australia’s real Gross Domestic Product per capita), yield (airline yields are used as a proxy of average airline fares and are based on Australia’s real best discount economy air fares), Australia’s unemployment numbers, recorded bed capacities at Australia’s tourist accommodation establishments\(^1\), and two dummy variables. The first dummy variable explained the impact of the evolving Virgin Australia business model from a low cost carrier model to a full service network carrier\(^2\) (FSNC) on Australia’s low cost carrier traffic. Australia’s low cost carriers’ traffic in Australia has decreased significantly since 2011 primarily due to this transition in Virgin Australia’s business model evolution. Thus, the dummy variable reflecting the Virgin Australia changing business model (DUMMY 1) is then introduced as a zero for the period from Quarter 1 (2002) to Quarter 4 (2010) and one from Quarter 1 (2011) to Quarter 2 (2012).

The second dummy variable accounted for the loss of capacity following the collapse of Ansett Australia. At the time of its collapse in 2001, Ansett Australia’s domestic Australian market share was 35 per cent (Virgin Blue held around 10 per cent and Qantas had a 55 per cent market share) (Prideaux 2003). Ansett Australia experienced financial problems and was placed into receivership on September 14, 2001 (Easdown, Wilms 2002). The collapse of Ansett Australia had a major impact on the tourism industry, especially in regional areas where Ansett’s subsidiaries provided substantial capacity. Whilst the other incumbent airlines increased seating capacity, the demand for seats exceeded supply for several months (Prideaux 2003). Australia’s GDP and air fares are expressed in real terms to remove the effect of inflation (Abed et al. 2001; Ba-Fail et al. 2000). Figure 1 presents the study’s target time series, and shows Australia’s annual low cost carrier enplaned passengers and RPKs.

Multiple regression analysis was carried out using the Statistical Package for Social Science (SPSS) Version 22 program.

The two models, PAX and RPKs, offering the best fit for Australia’s low cost carrier traffic demand are: (1):

\[ \ln PAX = -64.09 - 0.40 \ln x_1 + 1.42 \ln x_2 - 0.48 \ln x_3 + 4.80 \ln x_4 - 0.72 x_5 - 0.14 x_6 \]

\( (\text{Adj.} R^2 = 0.969, \text{S.E.} = 0.065) \), and (2):

\[ \ln RPK = -81.60 - 0.40 \ln x_1 + 1.58 \ln x_2 - 0.61 \ln x_3 + 6.05 \ln x_4 - 0.65 x_5 - 0.29 x_6 \]

\( (\text{Adj.} R^2 = 0.956, \text{S.E.} = 0.091) \).

---

1 Based on Australian tourist accommodation establishments with 15 rooms or more.

2 A "full service network carrier" (FSNC) (or "legacy carrier") is an airline that focuses on providing an extensive range of both pre-flight and on-board services, including different service classes, and connecting flights (Ehmer et al. 2008).
Here PAX stands for passengers enplaned by Australia’s domestic LCCs, RPK stands for revenue passenger kilometres performed by Australia’s domestic LCCs, \( x_1 \) is the airfare, \( x_2 \) is Australia’s real GDP per capita, \( x_3 \) is Australia’s unemployment size, \( x_4 \) is the recorded bed capacities at Australia’s tourist accommodation establishments, \( x_5 \) is the dummy variable (Dummy 1) reflecting Virgin Australia’s changing business model and \( x_6 \) is the dummy variable (Dummy 2) accounting for the loss of capacity following the collapse of Ansett Australia.

The modelling results are summarised in Table 1 (least square coefficients with standard errors and absolute t-values in parentheses) and shows a high \( R^2 \) for both models, which indicates that the variations between the variables are explained well by the models both of which have very small standard errors. High “\( t \)” values for the coefficients indicate that these variables are stable. The minimum squared error (MSE) of the estimated PAX and RPKs models with respect to the sample over Quarter 1 2002 to Quarter 2 2012 are 0.004 and 0.007, respectively. Attempts to use additional variables failed to improve the models’ fit (Alekseev, Seixas 2009). Also, a multicollinearity problem was found when including Australia’s real GDP and Australia’s population size as independent variables in the same model.

### 2.2. The artificial neural network modelling

#### 2.2.1. Artificial neural network architecture

Artificial neural networks (ANNs) are a method of using computer software to learn to recognise patterns in given data (BaFail 2004; Omidvar, Elliott 1997). ANNs capture the inherent information from a considered set of variables and learn from the existing data, even when noise is present (Garrido et al. 2014; Kasabov 1996). Hence, no formulation or a priori model is required (Watts, Worner 2008). A neural network can be trained to perform a particular function by adjusting the values of connections (weights) between elements (Kunt et al. 2011). During the training process, the ANN is able to detect complex relationships between the input and output data and perform synthesis (Sineglazov et al. 2013). Once the ANN has been trained on the sample of the given data-set, it can make estimations through the detection of similar patterns in future data (BaFail 2004).

### Table 1. The determinants of Australia’s LCC air travel demand

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>PAX</th>
<th>RPKs</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 ) Air fare</td>
<td>-0.40</td>
<td>-0.40</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.10)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>( t )</td>
<td>(-3.98)</td>
<td>(-2.85)</td>
</tr>
<tr>
<td>( x_2 ) GDP per capita</td>
<td>1.42</td>
<td>1.58</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.32)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>( t )</td>
<td>(4.47)</td>
<td>(3.54)</td>
</tr>
<tr>
<td>( x_3 ) Unemployment</td>
<td>-0.48</td>
<td>-0.61</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.12)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>( t )</td>
<td>(-4.09)</td>
<td>(-3.64)</td>
</tr>
<tr>
<td>( x_4 ) Bed spaces</td>
<td>4.80</td>
<td>6.05</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.59)</td>
<td>(0.82)</td>
</tr>
<tr>
<td>( t )</td>
<td>(8.20)</td>
<td>(7.37)</td>
</tr>
<tr>
<td>( x_5 ) Dummy 1</td>
<td>-0.72</td>
<td>-0.65</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.03)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>( t )</td>
<td>(-20.86)</td>
<td>(-13.40)</td>
</tr>
<tr>
<td>( x_6 ) Dummy 2</td>
<td>-0.14</td>
<td>-0.29</td>
</tr>
<tr>
<td>( SE )</td>
<td>(0.06)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>( t )</td>
<td>(-2.45)</td>
<td>(-3.69)</td>
</tr>
<tr>
<td>Constant</td>
<td>-64.09</td>
<td>-81.60</td>
</tr>
<tr>
<td>Adjusted ( R^2 )</td>
<td>0.969</td>
<td>0.956</td>
</tr>
<tr>
<td>( N )</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>( F )</td>
<td>215.52</td>
<td>151.07</td>
</tr>
<tr>
<td>Durbin Watson (DW)</td>
<td>1.49</td>
<td>1.38</td>
</tr>
</tbody>
</table>

ANNs have the ability to detect similarities in inputs, despite a particular input not ever being seen previously. This property provides ANNs with excellent interpolation capabilities, particularly when the input data may not be exact, that is, noisy (BaFail 2004).

The most general form of a neural network used in forecasting is shown in the following equation:

\[
Y = F [H_1(x), H_2(x), ..., H_n(x)] + u, \tag{2}
\]

where, \( Y \) is a dependent variable, \( X \) is a set of explanatory variables, \( F \) and \( H \)'s are network functions, and \( u \) is a model error term (BaFail 2004: 103).

The artificial neural network model is characterized by a network of three layers: input, output and hidden layers, which resembles the human body’s neural network (Lahmiri 2011; Mehrotra et al. 2000). Neural networks consist of a large number of simple processing elements called neurons which are organized into several layers and interconnected with each other through...
The factors that influence air travel demand are complex (Doganas 2009; Vasigh et al. 2008). Each factor is composed of elements that can stimulate or reduce air travel growth. For passenger air traffic demand forecasting purposes, these factors are more conveniently categorized into two broad groups: those external to the airline industry and those within the airline industry itself (Ba-Fail et al. 2000). Real GDP and real GDP per capita were used to measure the effect of income on Australia’s domestic LCC air travel demand.

The impact of Australia’s demographic changes was considered through Australia’s population and the number of unemployed persons (Tsekiris 2009; Young, Wells 2011). Population has a direct effect on the size of an air travel market and may cause a bias in the estimates if omitted. For instance, a large increase in air traffic may reflect a sudden increase in population rather than other effects (International Air Transport Association 2008). Unemployment rates have also been reported as being a determinant of air travel demand (Clark et al. 2009; Wensveen 2011). Ceteris Paribas, increasing levels of employment tend to increase air travel demand. Conversely, unemployment tends to depress air travel demand (McKnight 2010).

A decrease in the real cost of air travel also positively influences air traffic growth (Hanlon 2007; Holloway 2008). However, the measurement of the price of air travel is normally complicated by the presence of different fare classes offered by airlines (International Civil Aviation Organization 2006). Hence, airline passenger yields are often used as a proxy for air fares, which can be difficult to obtain given the wide use of a variety of, and fluctuating number of, discount fares. In the absence of changes in other factors influencing air travel demand, falling yields will tend to increase traffic. Conversely, rising yields will tend to reduce traffic volumes, subject to demand elasticities (Doganas 2009).

The growing demand for air travel, associated with the rapid growth of LCCs, has assisted tourism growth
(Davidson, Ryley 2010). The majority of the demand for LCC services is from leisure travellers (Graham 2006), although Mason (2005) has observed that there is an increasing number of business travellers who view LCC flights as providing good value, flexible ticket options, particularly on specific routes where frequency is tailored to satisfy business demand. Low-cost carriers are significant for the development of weekend, city or short-break tourism and are influencing the expansion of potential destinations (Graham, Shaw 2008). Visiting friends and relatives (VFR) traffic have also fed the LCCs (Bieger, Wittmer 2006). The LCCs are, therefore, extending the range of motivations and frequency of travel for private leisure reasons also by the use of their highly efficient websites, where customers can purchase not only a flight ticket, but also book a hotel, hire a car, and buy travel insurance (Olipra 2012). Thus, another explanatory variable included in the study relates to tourism attractiveness, which is expressed in terms of the tourist accommodation infrastructure, that is the reported bed capacity (Tsekeris 2009).

There are also a variety of other factors that may influence air travel demand, for example, jet fuel prices (Gesell 1993). Sharp increases in world oil prices have had important impacts on world air travel demand. In addition to the adverse impact on the global economy, airlines are often forced to increase air fares to cover the higher fuel costs, which often have a detrimental impact on air travel demand. This is because increases in oil prices result in higher air fares and therefore make leisure travel more expensive (Li 2010).

Furthermore, interest rates influence the balance between expenditure and saving (Cook 2007). High interest rates will inhibit economic activity, which can have a dampening effect on airline traffic (Wensveen 2011).

Four dummy variables were included in the modeling to control for the influence of the change in the business model of Virgin Australia from an LCC to a full service network carrier (FSNC), the Global Financial Crisis (GFC) during the period 2007 to 2009, the collapse of Ansett Australia in September 2001 which led to a temporary supply shortage until the second quarter of 2002, and the Commonwealth Games held in Melbourne from 15 to 26 of March 2006.

The availability of a consistent data set allows the use of quarterly data for the period 2002 to 2012. The data used in the ANN models were sourced from a variety of sources. Data on Australia’s real GDP and real GDP per capita, Australia’s unemployment numbers, population size and recorded bed capacities at Australia’s tourist accommodation establishments are from the Australia Bureau of Statistics (ABS). Australia’s interest rates are from the Reserve Bank of Australia (RBA). The airfare data are from the Bureau of Infrastructure, Transport and Regional Economics (BITRE). The data on Australia’s LCC domestic enplaned passengers and revenue passenger kilometres performed (RPKs) are from the Bureau of Infrastructure, Transport and Regional Economics (BITRE), Qantas Group (Qantas Airways 2009, 2013), Tiger Airways (Tiger Airways 2012a–2013g), Virgin Australia (Virgin Australia 2011, 2012, 2013) and Virgin Blue (Virgin Blue 2004–2009; Virgin Blue Holdings 2010) annual reports, and websites. World jet fuel prices (expressed in Australian dollars) were sourced from the US Energy Information Administration (EIA). To convert collected data from current prices to real or constant prices, the consumer price index at 2011 constant prices was used (Ba-Fail et al. 2000).

Twelve variables were considered as input variables in the two neural network models: Australia’s real GDP, Australia’s real GDP per capita, real best discount economy air fares, population size, unemployment numbers, Australia’s tourist accommodation establishment recorded bed capacities, world jet fuel prices, real interest rate, and four dummy variables controlling for the influence of the evolving Virgin Australia business model from an LCC to a FSNC, the GFC, Ansett Australia’s collapse and the Commonwealth Games in Melbourne in March 2006.

2.2.3. Model evaluation

Goodness-of-fit (GOF) statistics are useful when comparing results across multiple studies, for examining competing models in a single study, and also for providing feedback on the level of knowledge about the uncertainty involved in the phenomenon of interest. Five measures were used in the present study: mean square error (MSE), the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and correlation coefficient (R) (Kunt et al. 2011; Ruiz-Aguilar et al. 2014; Tiryaki, Aydın 2014).

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

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

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

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

\[
R = \frac{\sum_{i=1}^{N} (t_i - \bar{T})(t_i - \bar{d})}{\sqrt{\sum_{i=1}^{N} (t_i - \bar{T})^2 \sum_{i=1}^{N} (t_i - \bar{d})^2}},
\]

where \(t_i\) is the measured values, \(\hat{t}_i\) is the predicted values, \(N\) is the total number of data, and \(\bar{d}\) is the average of the predicted values (Tiryaki, Aydın 2014: 104).
2.2.4. Training and testing of artificial neural networks

Training is the algorithmic process in the hidden neuron where parameter weights can be adjusted appropriately to forecast accurately. Among various training algorithms, the back-propagation is the most popular algorithm used (Ba-Fail 2004; Claveria, Torra 2014; Faraway, Chatfield 1998; Zhang 2004). The basic way of using it is to propagate a gradient of the transfer function back and compare the actual output from output units with a target output, then re-adjust weights backwards in the network. Weights are adjusted and repeated until the mean squared error (MSE) between network prediction and actual data is close to the target (Jung, Wang 2007; Tiryaki, Aydn 2014; Zhang 2004).

For the purpose of the training process, neural networks are separated into three data sets: training is used for model fitting and selection, testing is used for evaluating the model's forecasting ability and validation data sets to determine the end point for the training process to avoid model over fitting (Alekseev, Seixas 2009; Garrido et al. 2014; Tiryaki, Aydn 2014; Zhang et al. 1998). Indeed, over-fitting is a major concern with neural network model building (Remus, O'Connor 2001; Smith, Ragsdale 2010), as it can lead to predictions that are beyond the range of the training data (Jeon 2007). In order to avoid over-fitting the study's models, the study's neural network design was carried out using three data sets: training, validation, and testing, which were randomly divided into a 70:15:15 ratio (Garrido et al. 2014; Kunt et al. 2011; Tiryaki, Aydn 2014). Importantly, a cross validation process has to be carried out during the training phase to avoid over-fitting (Efendigil et al. 2009).

The objective of training is to minimize the global error, such as root mean square error (RMSE), mean average error (MAE), mean square error (MSE), and mean absolute percent error (MAPE). ANNs usually commence with randomized weights for all their neurons. This means that they do not know everything and, therefore, require training to solve a particular problem for which they are intended. When a satisfactory level of performance is reached, training is concluded and the network uses these weights to decide (Akgüngör, Doğan 2009).

The training set was used to adapt the synaptic weights of the multilayer network, utilising the back propagation of estimation errors (Haykin 1999). All inputs were inserted into the model and the networks trained. During the supervised learning process, an error function is defined. The synaptic weight values are iteratively updated until the provided output is as desired, and the error function descends along the surface towards a local minimum. In this study the training process stopped when it reached 1,000 epochs or 0.01 error tolerance (Efendigil et al. 2009).

To conclude the training phase, a validation data set was used. The stopping criterion was the mean square error (MSE) of the estimated demand with respect to the samples belonging to the validation set. The validation set was not used in adapting the weight vectors of the neural estimator, and was therefore able to detect over-fitting in the training phase (Alekseev, Seixas 2009).

For estimating the generalization capacity of the ANN forecasting model, a testing set was also used (Alekseev, Seixas 2009). Thus, after the training process was completed, a testing process was applied to ensure the model accuracy was sufficiently reliable. Once the values of the training set were determined, a data testing set was fed into the model and the output compared to the target value. The model was accepted if the difference was low enough (Garrido et al. 2014). The testing set simulates the forecasting of the samples (Alekseev, Seixas 2009). The neural network process is summarized in Figure 3.

![Fig. 3. The study's artificial neural network modelling process (adapted from Jiang et al. 2004)](image-url)
The most frequently used transfer function is the sigmoid or logistic function as it possesses favourable mathematical properties such as monotonicity, continuity, and differentiability, which are all important when training a neural network with gradient descent (Priddy, Keller 2005). An activation function is used as a boundary of output. These boundaries normally change from zero to one [0, 1] or from minus one to plus one [–1, +1] according to the type of activation function used in the ANN (Akgüngör, Doğan 2009).

This study used the sigmoid function in the hidden layer and the linear transfer function in the output layer. The Levenberg–Marquardt back propagation algorithm was used as the training algorithm since its convergence is stable and fast (Ruiz-Aguilar et al. 2014). The Neural Network tool box 8.0 within the framework of MATLAB R2012b (The MathWorks, Inc., USA.) software was used for modelling and simulation purposes.

2.2.6. Artificial neural modelling results

Two different ANN models were developed to predict air travel demand of Australia's domestic LCCs. The MLP model consisted of three layers having weight matrix \( W \), bias vector \( b \) and output vector \( p^i \) where \( i > 1 \). Figure 4 presents the optimum MLP model for predicting enplaned passenger traffic of Australia's low cost carriers and RPKs. The number of each layer is shown as a superscript to the variable of interest. Following Kunt et al. (2011), superscripts were used for identifying the source (second index) and destination (first index) for the various weights and other elements of the network.

![Fig. 4. The structure of the final Multi-Layer Perceptron neural network model (adapted from Kunt et al. 2011: 358)](image)

The weight matrix connected to input vector \( p^1 \) was labelled as input weight matrix \( IW^{1×l_1} \). The elements of layer 1, such as its bias, net input and output have superscript 1 to indicate that they were associated with the first layer (Kunt et al. 2011).

The matrices of layer weight (LW) and input weight (IW) were utilised in the MLP model. Data were randomly divided into three parts: training, testing, and validation (Alekseev, Seixas 2009; Kunt et al. 2011). The MLP model had 11 inputs, 9 neurons in the hidden layers and 1 neuron in the output layer. The output layer of the MLP model consisted of one neuron representing Australia’s LCCs domestic enplaned passengers or RPKs' values, respectively. As noted earlier, 70 per cent of the data were used in the training phase. Validation and testing data sets each contained 15 per cent of the original data.

Constant input 1 was fed to the bias of each neuron. The outputs of each intermediate layer were the inputs to the subsequent layer. Hence, layer 2 can be analysed as one-layer having 9 inputs, 1 neuron and 1 X 9 weight matrix \( W^2 \). The layer can be treated as a single-layer network in its own right. The layers of a MLP play different roles in the prediction process (Kunt et al. 2011). The back propagation algorithm was applied to determine errors and modification for the weight of the hidden layer neurons (Akgüngör, Doğan 2009). In this study, \( p^3 \) was the network output of interest and has been labelled as \( y \) (Rumelhart et al. 1986).

The objective of this network is to reduce error \( e \), which is the difference between \( t \) and \( p^i \) in which \( i > 1 \) and \( t \) is the target vector. The perceptron learning rule calculates desired changes (target output) in the weights and biases of the perceptron, given input vector \( p^1 \) and the associated error \( e \). Accordingly, the Least Mean Square Error (LMS) algorithm adjusts the weights and biases of the linear network so as to minimize this mean square error (Kunt et al. 2011).

The error at output neuron \( j \) at iteration \( t \) can be calculated by the difference between the desired output (target output) and the corresponding real output, \( e_j(t) = d_j(t) - y_j(t) \) So, Eq. (9) is the total error energy of all output neurons.

\[
e(t) = \frac{1}{2} \sum_{j \in c} e_j^2(t). \tag{9}
\]

Referring to Figure 4, the output of the \( k \)-th neuron in the \( l \)-th layer can be calculated by Eq. (10) in which \( f_2 = \log \text{sig} \) and \( f_3 = \text{purelin}:
\]

\[
y^l_k = f_k \left( \sum_{j=1}^{n_l} w_{jk}^l \cdot y_j^{l-1} \right), \tag{10}
\]

where \( 1 \leq l \leq 3 \), \( n_l \) refers to the number of neurons in layer \( l \). For the input layer the following holds \( l = 1 \), \( y^1_j = x_j \), as for the output layer \( l = 3 \), \( y^3_j = y_j \).

The mean square error (MSE) of the output can be computed by:

\[
E = \frac{1}{2} \left( d_j - y_j \right)^2 = \frac{1}{2} \left[ d_j - f_3 \left( \sum_{i=1}^{9} w_{ij}^3 \cdot y_j^2 \right) \right]^2. \tag{11}
\]

The steepest descent of MSE can be used to update weights by Eq. (12) (Yeung et al. 2010):

\[
w_{ij}^3(t+1) = w_{ij}^3(t) - \eta \frac{\partial E}{\partial w_{ij}^3}. \tag{12}
\]

The MSE performance index for the ANN is a quadratic function as shown in Eq. (11). Hence, the performance
index will either have one global minimum, a weak minimum or no minimum, depending upon the characteristics of input vectors (Kunt et al. 2011). Specifically, the characteristics of input vectors determine whether or not a unique solution exists (Hagan et al. 1996).

The performance of the ANN can be increased if relevant information is extracted for feeding the network. Here, the ANN can evaluate correlations of such intelligent variables in the original input data space. Considering all the information that was available and following the extensive review of the literature on the determinants of air transport demand, additional input variables that were previously identified but not used in the econometric analysis due to statistical insignificance were considered for feeding the input nodes of the neural estimator in conjunction with the existing variables (Alekseev, Seixas 2009). Additional vector components included: world jet fuel prices, Australia’s interest rates, plus 2 dummy variables, reflecting the Commonwealth Games in 2006, and the effects of the GFC.

The results of the PAX and RPKs MLP models are presented in Table 2 in the form of a prediction table. Table 2 shows the prediction level of Australia’s LCCs demand (as measured by RPKs and enplaned passengers, respectively) during training, testing, and validation phases.

Table 2. Prediction of the MLP models

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (PAX)</th>
<th>Model 2 (RPKs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td>Validation</td>
<td>0.990</td>
<td>0.998</td>
</tr>
<tr>
<td>Test</td>
<td>0.992</td>
<td>0.991</td>
</tr>
<tr>
<td>All</td>
<td>0.996</td>
<td>0.998</td>
</tr>
</tbody>
</table>

Figure 5 shows regression plots of the PAX model output with respect to training, validation and testing data. The value of the correlation coefficient (R) for each phase was also calculated (Kunt et al. 2011). The R value was around 0.996 for the total response in the MLP model. The solid lines in Figure 5 show a perfect linear fit between actual values and estimated values of passengers enplaned by Australia’s domestic LCCs. The correlation coefficient (R) between actual values and estimated values is another important indicator to check the validity of the model. Importantly, when the R value is close to 1, forecasting accuracy increases (Tiryaki, Aydin 2014).

The relationship between actual values and estimated values obtained in the RPKs model is shown in Figure 6. The R value was around 0.998 for the total response in the MLP model.

Training errors, validation errors and testing errors were plotted to determine validation errors in the training phase for both the PAX and RPKs models (Fig. 7). The best validation performance in the PAX model occurred at epoch 6 with MSE at 0.0036 (Fig. 7). The plot in Figure 7 shows the mean squared error commencing at a large value and decreasing to a smaller value, which indicates that network learning is improving. The plot in Figure 7 has three lines, because 42 input and target vectors were randomly divided into three sets (Garrido et al. 2014; Kunt et al. 2011). 70 per cent of the vectors were used for training the network. 15 per cent of the vectors were used for validating how well the network model was generalised. The training of vectors continues for as long as it takes for training to reduce the network error on validation vectors. After the network has memorized...
the training set, training concludes. This technique automatically avoids the problem of over-fitting the model, which plagues many optimization and learning algorithms (Kunt et al. 2011). As previously noted, the training process stopped when it reached 1,000 epochs or 0.01 error tolerance (Efendigil et al. 2009).

In order to estimate the generalization capacity of the neural PAX and RPKs forecasting models, a testing set compromising the remaining 15 per cent of the vectors was used. This set was only presented to the neural estimator following conclusion of the training; hence, it did not participate in the training phase (Alekseev, Seixas 2009).

Fig. 7. The validation error in the PAX model

The RPKs model’s training errors, validation errors and testing errors were also plotted to find the validation error in the training phase. The best validation performance in the model occurred at epoch 12 with MSE at 0.0013 (Fig. 8). Similar to Figure 7 (PAX model), the plot shows a decrease in the MSE of the network which indicates that network learning is improving.

Fig. 8. The validation error in the RPKs model

Australia’s actual LCC and estimated enplaned passengers during Quarter 1 (2002) to Quarter 2 (2012) are plotted and shown in Figure 9.

Fig. 9. A comparison of Australia’s actual and estimated LCCs enplaned passengers

Australia’s LCCs actual domestic revenue passenger kilometres (RPKs) and estimated RPKs from Quarter 1 (2002) to Quarter 2 (2012) are plotted and shown in Figure 10, indicating the accuracy of the estimations.

Fig. 10. A comparison of Australia’s LCCs actual and estimated RPKs

3. Comparison of classical and neural network models

This study used a classical linear regression model and an artificial neural network (ANN) to predict Australia’s LCCs domestic air travel demand. The final results showed that the ANN offered improved estimation over the linear regression models. Table 3 and Table 4 present root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) of the estimated models. These results suggest that the constructed ANNs are promising for modelling Australia’s LCCs air travel demand.

Table 3. The final results of the objective function in Australia’s LCC PAX models

<table>
<thead>
<tr>
<th>Error</th>
<th>Classical model</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.049</td>
<td>0.018</td>
</tr>
<tr>
<td>MSE</td>
<td>0.004</td>
<td>0.001</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.60%</td>
<td>0.21%</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.062</td>
<td>0.034</td>
</tr>
</tbody>
</table>
Table 4. The final results of the objective function in Australia’s LCC RPKs models

<table>
<thead>
<tr>
<th>Error</th>
<th>Classical model</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.067</td>
<td>0.016</td>
</tr>
<tr>
<td>MSE</td>
<td>0.007</td>
<td>0.001</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.80%</td>
<td>0.19%</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.086</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Figures 11 and 12 compare the actual output values of Australia’s LCC domestic enplaned passengers and revenue passenger kilometres performed (RPKs) with the predicted values of the models tested in the study. This graphical presentation highlights a considerable overlap between the real and predicted outputs from the MLR and ANN models, indicating that the models successfully predict Australia’s LCC domestic demand with a high level of accuracy.

4. Conclusions

Forecasting is regarded as one of the most critical areas of airline management. This study developed models for predicting Australia’s low cost carriers’ domestic passenger demand, using enplaned passengers and revenue passenger kilometres (RPKs), as the measures of airline traffic demand. Two approaches were compared; classical regression modelling and artificial neural network modelling. Classical linear regression econometric models were developed and the statistical relationship between key demand-influencing factors and the corresponding level of Australia’s low cost carrier passenger traffic were tested. The statistical measures for evaluating the models show that the following models are found to be the most appropriate models to predict Australia’s domestic LCCs air travel demand:

**Domestic air travel demand (PAX)**

\[
Y_1 = -64.09 - 0.40lnx_1 + 1.42lnx_2 - 0.48lnx_3 + 4.80lnx_4 - 0.72x_5 - 0.14x_6
\]

**Domestic air travel demand (RPKs)**

\[
Y_2 = -81.60 - 0.40lnx_1 + 1.58lnx_2 - 0.61lnx_3 + 6.05lnx_4 - 0.65x_5 - 0.29x_6,
\]

where:

- \(Y_1\) is the number of Australia’s LCC passengers;
- \(Y_2\) is Australia’s LCC RPKs;
- \(x_1\) is Australia’s real best economy discount airfare;
- \(x_2\) is Australia’s real GDP per capita;
- \(x_3\) is Australia’s unemployment size;
- \(x_4\) is recorded bed capacities at Australia’s tourist accommodation establishments;
- \(x_5\) is the dummy variable (Dummy 1) reflecting Virgin Australia’s changing business model;
- \(x_6\) is the dummy variable (Dummy 2) accounting for the loss of capacity following the collapse of Ansett Australia.

Both models are good in terms of goodness of fit measures and MAE, MSE, MAPE and RMSE.

The ANN was applied for training, testing and validation and contained eleven inputs and nine neurons in the hidden layer and one neuron in the output layer. 70% of the data was used in the training phase with the remaining data divided into validation (15%) and testing (15%). The R-value of Model 1 (PAX) was around 0.996 and Model 2 (RPKs) was 0.998, respectively.

The comparison of the modelling results shows that the performance of the ANNs was superior and offers improved estimation over the classical MLR models. The use of multiple modelling approaches in this study provided a greater understanding of the determinants of Australia’s LCCs passenger demand and allowed for higher prediction accuracy (ANN) than in the traditional MLR approach.

References


