APPLYING DATA MINING TECHNIQUES TO FORECAST NUMBER OF AIRLINE PASSENGERS IN SAUDI ARABIA (DOMESTIC AND INTERNATIONAL TRAVELS)

Abdullah Omer BaFail
King Abdul Aziz University
Jeddah, Saudi Arabia

ABSTRACT

This work involves forecasting the number of domestic and international airline passengers in Saudi Arabia. The method used is the neural network technique. Annual data 1975 to 1986 was used and categorized into 16 variables. The forecast was obtained using the Model Quest Miner package, which uses some historical data for developing the model then proceeds to an evaluation phase. The period used for developing the model for the number of passengers was 18 years, while the period used for evaluation was 6 years. Samples of output are presented for each model. Plots of forecasts versus the actual number are shown, together with the percentage of trend fit. Results indicated that the oil gross domestic product, population size and per capita income were found to be the most contributing variables that affect the number of passengers in the Saudi Arabian airline sectors.

INTRODUCTION

Fortunately the Kingdom of Saudi Arabia is endowed with numerous natural opportunities for air travel because of its geographical location, being the site of the two holy mosques, its vast land area spreading from Showrorah in the south to Tabouk in the north, its rapid development in all spheres of life and its friendly relations with world community.

Abdullah Omer BaFail graduated in Systems Engineering from King Fahad University of Petroleum and Minerals, Dammam, Saudi Arabia in 1979. He joined the Department of Industrial Engineering, King Abdul Aziz University Jeddah in 1979 as a Graduate Assistant. He took his Masters degree in Industrial and Systems Engineering at University of Colorado in 1982. In 1989 took his Ph.D. in Industrial and Systems Engineering from Arizona State University. He was promoted to the post of Assistant Professor in 1989 and then in 1997 as Associate Professor. Currently he is the Chairman of the Department of Industrial Engineering. He has to his credit more than 15 papers published in international and national journals and conferences. He has several ongoing consultancy projects with the Government and some major private industrial firms in Saudi Arabia.

Air transportation in the Kingdom has undergone considerable expansions and developments during the past years. There are 25 international and domestic airports. The number of passengers (arriving & departing) handled by all airports increased at an average annual rate of 15%, rising from 1.6 million in 1970 to 30 million in 1998.

The Saudi Arabian Airlines (Saudia) took the operational responsibility to run the air transportation business in the Kingdom, and has made a vital contribution to the development of the Kingdom (A Brief Outline, 1990; Update Forecast, 1979). Saudia is constantly endeavoring to innovate, and to plan ahead for improved service to customers. One of the improvements to the current business is forecasting and analyzing the air travel market.

The main objective of this paper is to develop a forecasting model to forecast the number of airline passengers (international and domestic) in five major cities in Saudi Arabia. The cities considered for the study are Riyadh, Jeddah, Madina, Dhahran, and Taif. The methodology follows artificial neural network, using the Model Quest Minor package. The next section explains air travel market analysis, followed by an analysis of the forecasting methodology, the data analysis and conclusions.

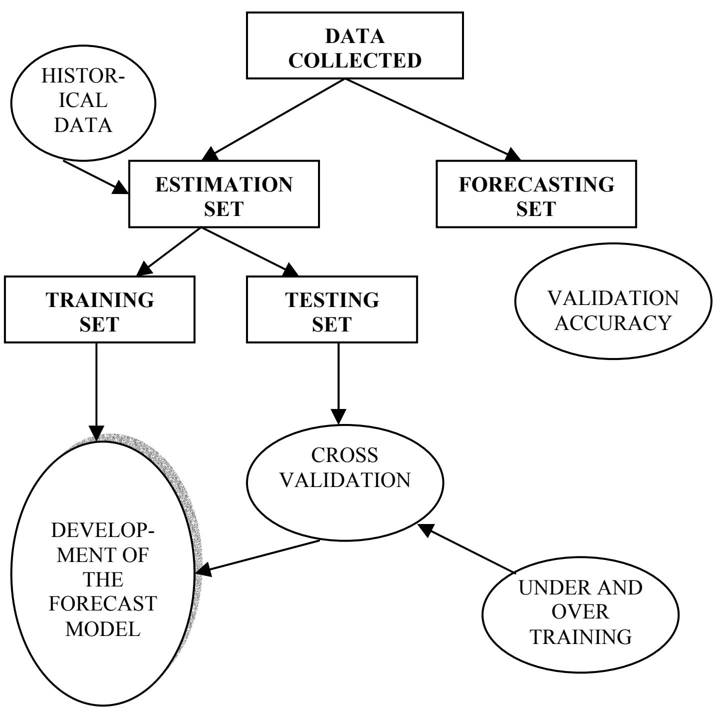
AIR TRAVEL MARKET ANALYSIS

Information relevant to the characteristics of the relationship among the air travel growth factors enabled the author to make a list of 16 variables which might influence the air travel demand for domestic and international travel (International, 1994). This list is as follows:

1. Oil Gross Domestic Product (Oil - GDP)
2. Private Non-Oil Gross Domestic Product (P-S)
3. Government Non-Oil Gross Domestic Product (G-S)
4. Total Non-Oil Gross Domestic Product (YND-G)
5. Total Gross Domestic Product (T-G)
6. Consumer Price Index (CPI)
7. Per Capita Income (PERC- IN)
8. Import of Goods and Services (IGS)
9. Exchange Rate (Saudi Riyals /U)
10. Population Size (P- Size)
11. Total Expenditures (T - Exp)
12. Private Consumption Expenditures (PFC-EXP)
13. Government Consumption Expenditures (GFC-EXP)
14. City Domestic (City-DOM)
15. City International (City- INT)
16. Year

The historical data in terms of real prices (in Million Saudi Riyals) of the above variables against the years covering the period 1975 to 1998 are published in a paper by Khalid Al-Bassam (1997). This historical data is called the estimation set. In this study, a fraction of the overall available data is reserved for validating the accuracy of the developed forecast model. This reserved data set is called the forecasting set because no information contained in it is used in any form during the development of the forecast model. The data in the forecasting set are used for testing the true extrapolative properties of the developed forecast model. The estimation set is further divided into a training set and a testing set. Information in the training set is used directly for the determination of the forecast model, whereas information in the testing set is used indirectly for the same purpose. In particular, information in the testing set is used for out-of-sample testing, that is, for cross-validation, allowing prevention of under-training (under-fitting) and over-training (over-fitting). Prevention of under-training and over-training is crucial in the development of a forecast model which extrapolates well beyond the estimation set. Figure 1 presents the model development schematically.

Figure 1. Schematic representation of forecast model development.



FORECASTING METHODOLOGY

By the simplest definition data mining is an attempt to find patterns in data. There are many different methodologies to data mining such as cluster analysis, decision trees, categorization analysis, visualization, time series analysis, hybrid approaches, linkage analysis, and neural network (Update Forecast, 1979). In this study, neural network architecture will be used.

Neural Network Architecture

A neural network is a software (or hardware) simulation of a biological brain (sometimes called an artificial neural network or ANN). The purpose of a neural network is to learn to recognize patterns in a given data. Once the neural network has been trained on samples of the given data, it can make predictions by detecting similar patterns in future data. Software that learns is truly artificial intelligence.

Neural networks are able to detect similarities in inputs, even though a particular input may never have been seen previously. This property allows for excellent interpolation capabilities, especially when the input data is noisy (not exact). Neural networks may be used as a direct substitute for autocorrelation, multivariable regression, linear regression, trigonometric and other regression techniques.

Neural networks are often pictured as layers of functional nodes. The most general form of a neural network model used in forecasting can be written as:

$$Y = F\,[H1\,(x),\ H2\,(x),\ \ldots,\ Hn\,(x)] + u$$

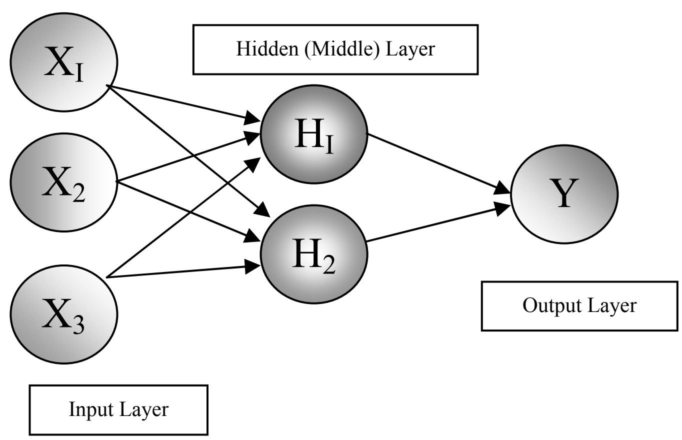
Where, Y is a dependent variable, X is a set of explanatory variables, F & H's are network functions, and u is a model error term.

Figure 2 shows a typical 3-layer back propagation network: an input layer, a hidden layer, and an output layer. Layer 1, the Xs, is called the input layer. The input layer is configured such that current and past values of the indicators (inputs) and of the forecasted variables (outputs) can be utilized (Barla, 1999). This input layer is connected to a hidden layer 2. Hs are the hidden layer activation functions. Each hidden layer has a collection of processing elements called nodes, which represents a different nonlinear function. Each node in a layer receives its input from the preceding layer through link elements called weights. The weights, along with the bias, are the free parameters of the model structure which must be adjusted using an appropriate learning algorithm and the information contained in the training set. Finally, layer 3, the Ys, is the output layer, which takes the intermediate results from the hidden layer and computes a prediction. In practice, the number of layers, number of nodes, and specific functional form in each layer must be determined by pragmatic experimentation.

Artificial Neural Network Learning Algorithm

For a given ANN architecture and a training set, the basic mechanism behind most supervised learning rules is the updating of the weights and the bias terms, until the mean squared error (MSE) between the output predicted by the network and the desired output (the target) is less than a pre-specified tolerance. The method used to compute these weights is the Adaptive Back Propagation (ABP) algorithm. First, random values are assigned to all the weights. Second, a set of training data, consisting of multidimensional input vectors with matching output values, are submitted to the model. For each input, the model output is compared to the known, correct output. Third, the weights are then iteratively adjusted to reduce the MSE over the entire training set. The iterative algorithm to adjust the weights is an application of the chain rule of calculus. Derivatives of the MSE are computed, and adjustments are made using a gradient search method that propagates back down the network from layer 3 to layer 2.

Figure 2. . Input and output layers of a typical 3-layer back propagation network.



While a neural network model can be programmed from scratch, the availability of neural network software packages makes the process easier. These packages supply predefined neural network architectures, like back propagation, and include algorithms to handle the delicate iterative computation of neural network weights.

In this paper the Model Quest Miner package is utilized. The Model Quest Miner package is a modeling tool that automatically learns numeric knowledge from the data. This model is implemented as a mathematical model called Statistical Networks. A Statistical Network is a network of functional nodes. Each node contains a mathematical function that computes an output given a number of inputs. Information in a statistical network flows from the input variables through the network to the output variable.

Steps in Developing a Model

There are four basic steps to the development of a neural network model:

1. Selection of input: the raw market data variables (closing price, open interest, volume, etc.).
2. Calculation of new variables: functions of the input in Step 1. Examples include changes in volume and variance of daily changes over the immediate past.
3. Reduction of input data dimensionality.
4. Selection of neural network architecture and parameters.

These steps are repeated until a suitable model is obtained. By comparing model output with observed patterns in the input data, model weaknesses are reserved which suggests options for step 2, the creation of new technical variables.

DATA ANALYSIS

The data available for the 16 input variables were used to obtain the forecast, using the Model Quest Miner package. The estimation set for the model consisted of 18 years of data, while the evaluation set consisted of 6 years of data. The samples of the output results are presented. In addition, plots of the forecasts versus the actual number of passengers (domestic and international) for the five cities are shown in figures 3 through 12.

City of Dhahran (Domestic Travel)

The following indicators were selected as the most influencing and chosen by the model for forecasting the number of domestic passengers in the city of Dhahran: oil gross domestic product lag[1] 6 years, private non-oil gross domestic product lag 1 year, Import of goods and services lag 10 years, and population size lag 2 years. Similarly the following indicators were used

---

[1]The term “lag” is used to indicate data that is a certain number of years before the specific time period modeled.

for forecasting the number of international passenger in the city of Dhahran: oil gross domestic product lag 16 years, per capital income lag 5 year, population size lag 4 years, and the Dhahran's domestic flights lag 3 years.

Table 1 and figures 3 and 4 show the domestic and international actual and forecasted number of passengers for the city of Dhahran for the years 1993 through 1998. Forecasts underestimated the actual travel. The Mean Absolute Percentage Error (MAPE) for domestic travel is about 10%, while for international travel is about 3%.

Table 1. Domestic and international actual and forecasted number of passengers, Dhahran, 1993-1998

	Domestic Travel			*International Travel*		
Year	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*
1993	1,713.598	1,676,565	2.16	1,228,885	1,205,433	1.9
1994	1,743,335	1,628,101	6.6	1,248,262	1,237,908	1.3
1955	1,853,441	1,588,099	14.3	1,115,282	1,079,365	3.1
1996	1,653,363	1,341,018	18.8	1,129,537	1,163,185	2.9
1997	1,984.035	1,864,543	6.1	1,172,357	1,118,611	4.5
1998	2,108,150	1,801,502	14.5	1,200,591	1,248,262	3.9

Figure 3. Domestic actual and forecasted number of passengers for the city of Dhahran for the years 1993 through 1998

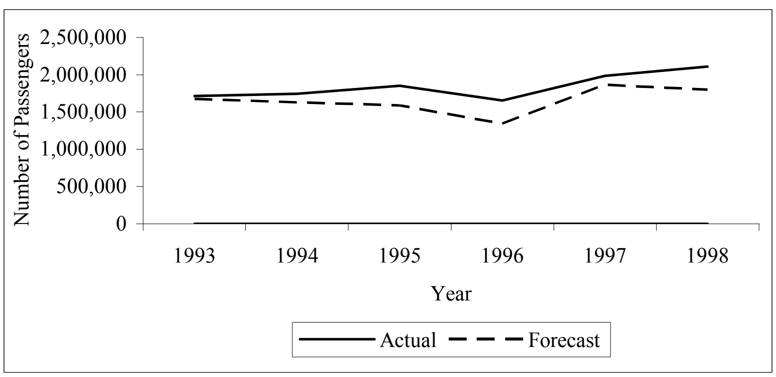
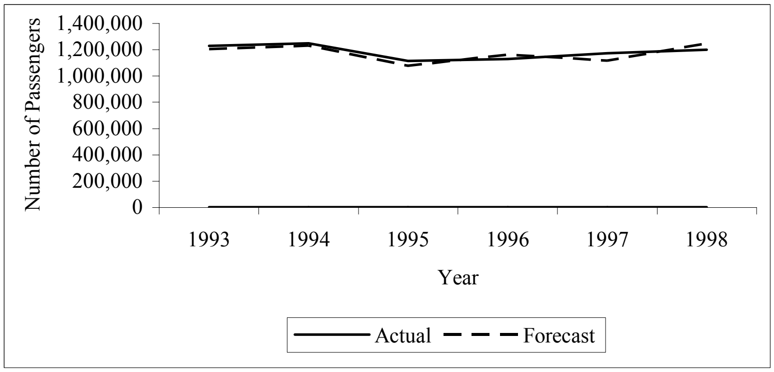


Figure 4. International actual and forecasted number of passengers for the city of Dhahran for the years 1993 through 1998



City of Madina

The following indicators were selected as the most influencing and chosen by the model for forecasting the number of domestic passengers in the city of Madina: oil gross domestic product lag 5 years, and per capita income lag 5 years. Similarly, the following indicators were used for forecasting the number of international passenger in the city of Madina: oil gross domestic product lag 10 years, per capita income lag 3 years, and Madina international flights lag 3 years.

Table 2 and figures 5 and 6 show the domestic and international actual and forecasted number of passengers for the city of Madina for the years 1993 through 1998. Forecasts underestimated the actual travel. The MAPE for domestic travel is about 8%, while for international travel is about 3%.

Table 2. Domestic and international actual and forecasted number of passengers, Madina, 1993-1998

	Domestic Travel			*International Travel*		
Year	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*
1993	1,005,669	985,960	1.9	142,379	140,245	1.4
1994	1,056,551	979,008	7.3	155,334	151,919	2.1
1955	1,074,002	1,008,526	6.1	118,066	114,247	3.2
1996	955,342	859,210	10.1	122,369	118,648	3.1
1997	1,146,411	1,057,704	7.8	131,178	104,811	2.1
1998	123,065	1,057,704	14.2	164,214	155,334	5.4

Figure 5. Domestic actual and forecasted number of passengers for the city of Madina for the years 1993 through 1998

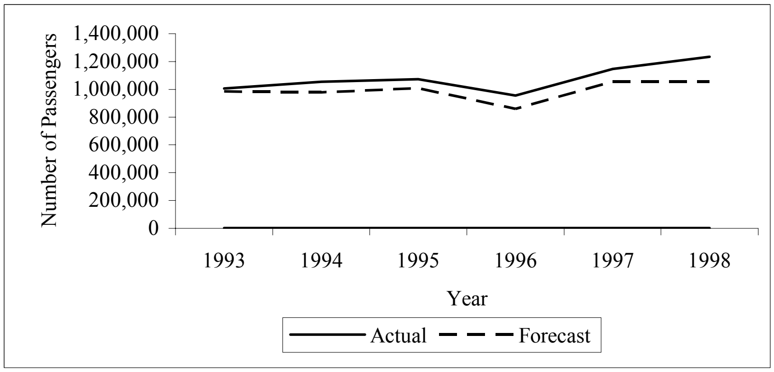
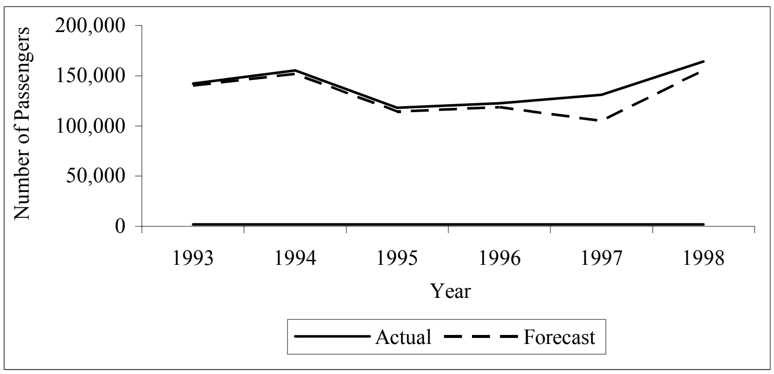


Figure 6. International actual and forecasted number of passengers for the city of Madina for the years 1993 through 1998



City of Riyadh

The following indicators were selected as the most influencing and chosen by the model for forecasting the number of domestic passenger in the city of Riyadh: oil gross domestic product lag 13 years, per capita income lag 3 years, and population size lag 2 years. Similarly, the following indicators were used for forecasting the number of international passenger in the city of Riyadh: oil gross domestic product lag 11 years, per capita income lag 1 years, total expenditures lag 2 years, and Riyadh domestic flights lag 4 years.

Table 3 and figures 7 and 8 show the domestic and international actual and forecasted number of passengers for the city of Riyadh for the years

1993 through 1998. Forecasts underestimated the actual travel. The MAPE for domestic travel is about 14%, while for international travel is about 7%.

Table 3. Domestic and international actual and forecasted number of passengers, Riyadh, 1993-1998

	Domestic Travel			*International Travel*		
Year	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*
1993	5,103,385	5,049,011	1.8	2,65,5847	2,751,197	3
1994	4,182,880	4,11,5535	1.6	2,711,855	2,690,474	7
1955	5,295,891	4,315,128	18.5	2,371,369	2,317,138	4
1996	4,880,131	4,593,008	20.0	2,428,567	2,341,402	4
1997	5,616,157	4,930,858	12.2	2,527,551	2,509,063	7
1998	5,926,781	4,063,034	31.0	2,731,190	23,07,566	15

Figure 7. Domestic actual and forecasted number of passengers for the city of Riyadh for the years 1993 through 1998

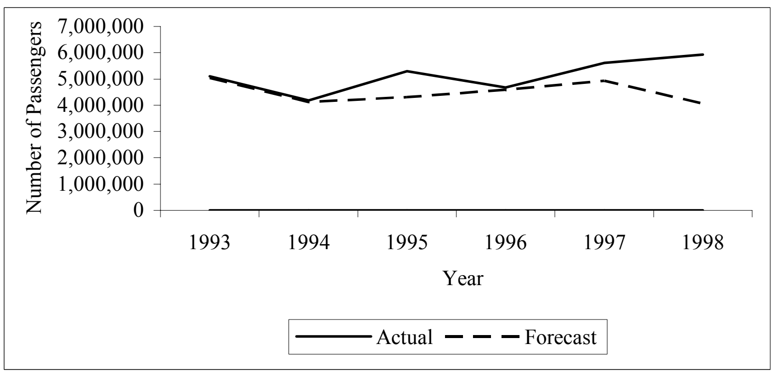
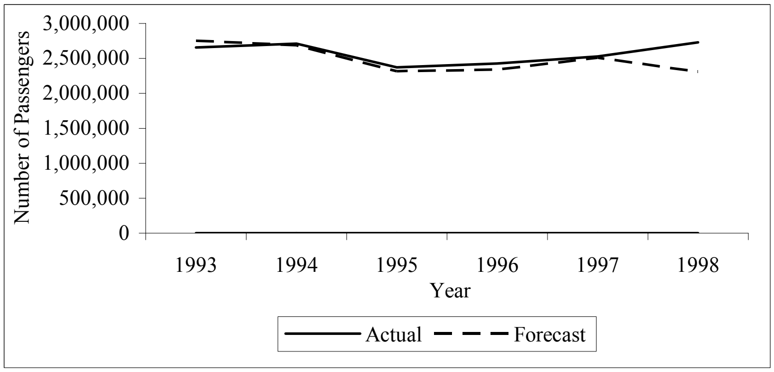


Figure 8. International actual and forecasted number of passengers for the city of Riyadh for the years 1993 through 1998



City of Jeddah

The following indicators were selected as the most influencing and chosen by the model for forecasting the number of domestic passenger in the city of Jeddah: oil gross domestic product lag 5 years, consumer price index lag 2 years, import of goods and services lag 1 year, and population size lag 3 years. Similarly, the following indicators were used for forecasting the number of international passenger in the city of Jeddah: oil gross domestic product lag 10 years and population size lag 3 years.

Table 4 and figures 9 and 10 show the domestic and international actual and forecasted number of passengers for the city of Jeddah for the years 1993 through 1998. Forecasts overestimated the actual travel. The MAPE for domestic travel is about 13%, while for international travel is about 4%.

Table 4. Domestic and international actual and forecasted number of passengers, Jeddah, 1993-1998

	Domestic Travel			*International Travel*		
Year	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*
1993	4,406,475	4,353,381	1.2	4,121,391	4,135,741	0.3
1994	4,459,569	4,225,227	5.2	4,431,179	4,365,286	1.5
1955	4,761,548	4,349,457	8.5	3,783,673	3,684,908	3.6
1996	4,263,140	4,406,475	8.3	3,840,078	3,716,900	3.2
1997	5,115,758	3,504,846	14.8	4,015,412	3,868,690	3.6
1998	5,428,940	3,504,846	26.7	4,20,0361	3,978,450	5.2

Figure 9. Domestic actual and forecasted number of passengers for the city of Jeddah for the years 1993 through 1998

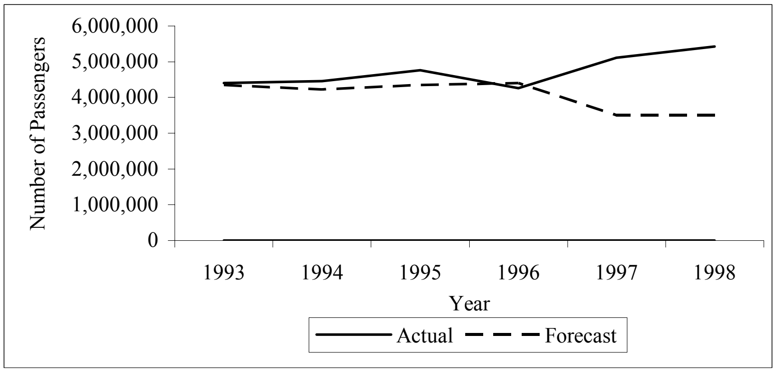
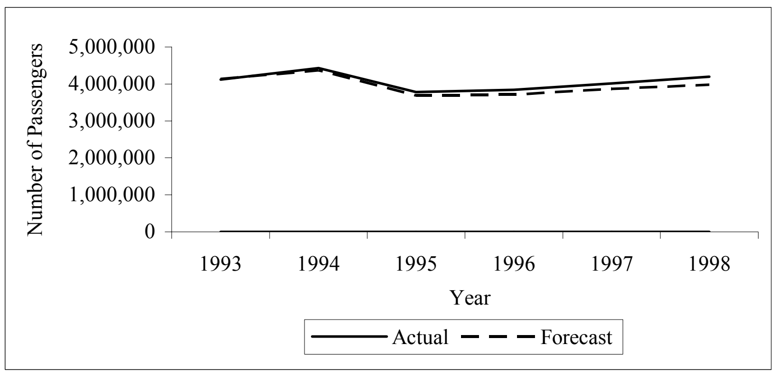


Figure 10. International actual and forecasted number of passengers for the city of Jeddah for the years 1993 through 1998



City of Taif

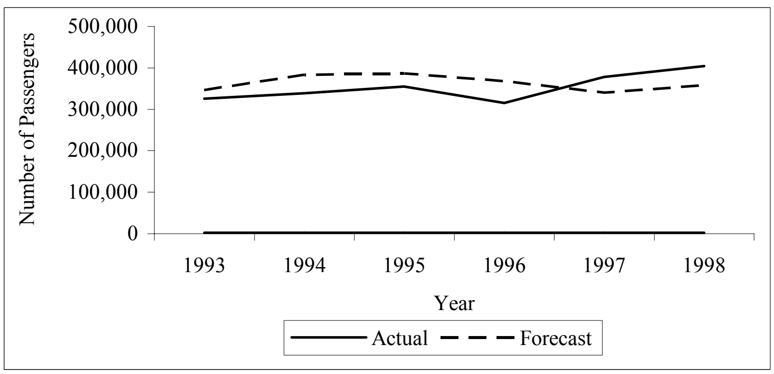
The following indicators were selected as the most influencing and chosen by the model for forecasting the number of domestic passenger in the city of Taif: oil gross domestic product lag 10 years, private consumption expenditures lag 2 years, and total gross domestic product lag 2 years. Similarly, the following indicators were used for forecasting the number of international passenger in the city of Taif: oil gross domestic product lag 10 years, and per capita income lag 5 years.

Table 5 and figures 11 and 12 show the domestic and international actual and forecasted number of passengers for the city of Taif for the years 1993-1998. Forecasts underestimated the actual domestic travel, and overestimated the international travel. The MAPE for domestic travel is about 13%, while for international travel is about 16%.

Table 5. Domestic and international actual and forecasted number of passengers, Taif, 1993-1998

	Domestic Travel			*International Travel*		
Year	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*	*Actual Number of Passengers*	*Forecast Number of Passengers*	*Mean Absolute Percentage Error*
1993	326,172	346,818	6.3	8,914	7,748	13
1994	338,948	383,295	13.0	8,729	7,486	14
1955	355,312	386,744	8.8	6,080	5,894	3
1996	315,374	368,615	16.8	6,552	5,174	21
1997	378,449	340,517	10.0	7,034	6,427	9
1998	404,181	358,564	11.2	7,612	5,894	22

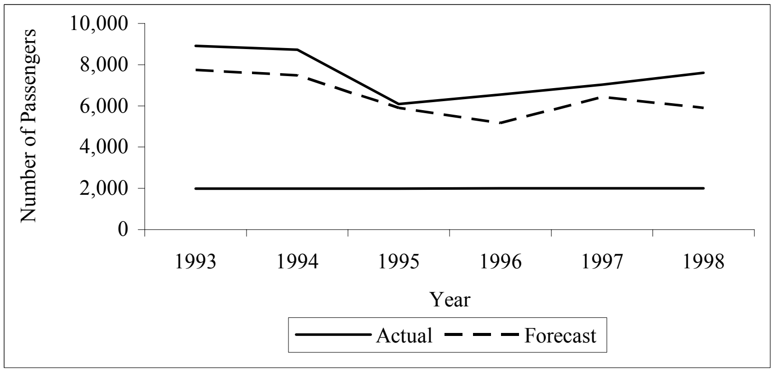
Figure 11. Domestic actual and forecasted number of passengers for the city of Taif for the years 1993-1998



RESULTS AND RECOMMENDATIONS

A suitable forecast model for the Saudi Airlines sector is essential for several reasons. In the first place, the demand is highly fluctuating and seasonal. The seasonality is chiefly due to Hajj and Umrah the two religious rituals among Muslims, which fall on Dil Hijja and Ramadan months. At each of these times, approximately 2 million pilgrims arrive in Saudi Arabia from various parts of the world. In the second place, the competition in the international sector of airline industry has caught up with the Saudi Arabian market, too, with the increased entry of foreign airline services.

Figure 12. International actual and forecasted number of passengers for the city of Taif for the years 1993-1998



Due to these reasons, results of the analysis were subjected to rigorous study. An overall examination of the data indicates that there was a general downward trend observed in both domestic and international sectors during the years 1995 to 1996. Saudi Arabian Airlines, which is the monopolist airline operators in the Kingdom, opened up several new routes beginning in 1993. Flights operated to Asmara, Orlando, Florida and several other places. This, together with the lack of new aircrafts, could have been the reason for the overall decline in the number of passengers during the 1995 to 1996 periods. However 61 new aircrafts were purchased from McDonnel and Boeing and a new corporate identity was launched in 1996. This boosted the passenger usage of the airlines.

Attempts were made to identify the crucial factors that influence airline usage. The indicators found to be most influential are shown in Table 6. It is seen that the most influential factors are oil gross domestic product and per capita income in the domestic and international sectors. Oil production is one key factor which influences most of the Saudi trading and commercial sectors. Therefore oil gross domestic product appearing as an indicator is quite understandable. Similarly population size is a deciding factor in the major cities like Dammam, Riyadh and Jeddah. The expatriate communities from India, Bangladesh and Pakistan in these major cities contribute much to the domestic as well as international services. In Jeddah the international sector forecast values closely follow the actual data. The most contributing figure in Jeddah sectors is definitely the floating population. Population size does not seem to be influencing the Madina and Taif sectors.

In view of the fluctuating nature of the passenger usage of airline services in Saudi Arabia, certain suggestions are made here. Most of these recommendations are in order to improve the resilience of the system to the fluctuations in demand and supply.

Table 6. Factors that influence airline usage, for domestic and international travel, 1993-1998

City	*Domestic*	*International*
Dhahran	Oil gross national product Per capital income Import of goods and services Population size	Oil gross national product Per capital income Population size Dhahran domestic product
Madina	Oil gross national product Per capital income	Oil gross national product Per capital income Madina international product
Riyadh	Oil gross national product Per capital income Population size	Oil gross national product Per capital income Total expenditures Riyadh domestic product
Jeddah	Oil gross national product Consumer price index Population size	Oil gross national product Population size
Taif	Oil gross national product Private consumption expenditures Total gross domestic product	Oil gross national product Per capita income

Privatization of the airlines services. International usage of airline services is increasing every day. From the trend lines of the forecast and actual values it can be seen that there is much difference appearing in the data sets of all cities towards the end of the study period. The actual number of passengers especially in the domestic sector is more than the predicted values in all cities.

Population size. Though population size does not appear as an indicator in the Madina sector, it is a fact that pilgrims flow to the city during Hajj and Umrah periods. However most of them use road transport instead of airlines. This is mostly due to difficulties in scheduling flights to the Madina sector. Similarly though Taif is one of the most attractive tourist summer resorts in the country, airlines services usage is limited in this sector also due to inadequate flight scheduling. Solutions to these problems can be found by implementing hubbing in these sectors (Barla, 1999). Hubbing provides the airlines with increased flexibility in adjusting their capacity allocations across markets as new information about demand conditions become available. Hubbing also pools passengers from several markets into the same plane. This helps the airlines service company to lower costs of operation drastically.

REFERENCES

A Brief Outline on Planning and Research Division at Saudia. (1990). Saudi: Corporate Planning Department.

Al-Bassam, Khalid. (1997). *An econometric model for air travel demand.* Unpublished master's thesis, King Abdul Aziz University, Jeddah, Saudi Arabia.

Barla, P. (1999). Demand uncertainty and airline network morphology with strategic interactions. *Cahiers de recherché, No. 9907,* Université Laval - Département d'économique.

International Marketing Data and Statistics 1993 (18th ed.). (1993). London: Euromonitor Publications.

Update Forecast and Planning Assumptions: New Riyadh International Airport. (1979, July). Riyadh: Saudi Arabian Bechtel Company.