DEMAND FORECASTING FOR DOMESTIC AIR TRANSPORTATION

IN TURKEY

BY

ORHAN SIVRIKAYA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

T.C.

OKAN UNIVERSITY

INSTITUTE OF SOCIAL SCIENCES

BUSINESS MANAGEMENT

MAY, 2013

D O K T O R A
T E Z O N A Y I

ÖĞRENCİNİN

Adı ve Soyadı	**:Orhan Sivrikaya**	Numarası	: **102001012**
Anabilim/Bilim Dalı	**:İşletme Anabilim Dalı**	Tez Savunma Tarihi	: **20.05.2013**
Danışman	**:Prof.Dr.Enar TUNÇ**	Tez Savunma Saati	: **19.00**

Tez Başlığı : Demand Forecasting For Domestic Air Transportation in Turkey

TEZ SAVUNMA SINAVI, Lisansüstü Öğretim Yönetmeliği'nin **42. Maddesi** uyarınca yapılmış, sorulan sorulara alınan cevaplar sonunda adayın tezinin ……………………..'ne OYBİRLİĞİ / OYÇOKLUĞUYLA karar verilmiştir.

JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
PROF.DR.ENAR TUNÇ		Kabul
PROF.DR.ÖZER ERTUNA	Özer	KABUL
PROF.DR.TARGAN ÜNAL		KABUL
DOÇ.DR.BAHAR SENNAROĞLU		KABUL
DOÇ.DR.ÖZALP VAYVAY		KABUL

YEDEK JÜRİ ÜYESİ	İMZA	KANAATİ (KABUL / RED / DÜZELTME)
DOÇ.DR.A.ZAFER ACAR		
DOÇ.DR.EKREM DUMAN		

An Abstract of a Dissertation Submitted to Okan University in Partial Fulfillment of the Requirements for the degree of Doctor of Philosophy

Demand Forecasting For Domestic Air Transportation in Turkey

By

Orhan Sivrikaya

May 2013

Airline industry relies strongly on air travel demand forecasting for developing the operating strategies, including the arrangements of destinations, routes, fleet planning and schedule structuring, and resource planning. This research utilizes artificial neural network (ANN) in comparison with multivariate semi-logarithmic regression (MLR) model to accurately forecast air travel demand between domestic cities in Turkey. The objective is to focus on the parallel assessment of artificial neural network and multivariate linear regression model's forecasting performance, so the same inputs and the same data were used for the development and comparison of the two approaches. Nodes in the input layer represent independent variables of the problem, including the geo-economics and airline specific factors used in semi-logarithmic model.

The input layer of the neural network with two hidden layers contains nine nodes; Airline count, log (population), log (price), log (bedding capacity), distance, transit, travel time, travel match and schedule consistency. The single node in the output layer of the neural network represents the log (number of passenger) for particular city pair. Airline annual data by city pair in the year 2011 were used to build the prediction models, and then the models were testified for their prediction performance by using annual data of the year 2010. Experimental results showed that using the novel model ANN to forecast domestic air travel demand outperforms the conventional model MLR. The model accuracy seems to be pretty satisfactory to generate the valuable data which can be a good reference for air carriers or government authorities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am especially grateful to Prof. Dr. Enar Tunç, from whom I have received countless support and encouragement that helped me very much to get through many difficult time. To me, he is not only the supportive advisor to research, but also a great mentor in life. Thanks for what he has done for me like a good friend.

I want to extend my appreciation to Prof. Dr. Targan Ünal and Prof. Dr. Özer Ertuna for their participation in my advisory committee as well as their valuable guidance in developing the dissertation.

I am also grateful to my family, to whom I am truly indebted. It is the continued support and love from my wife and kids that make me believe that life does not consists of only white and black, it is more colorful.

Thank you all who have already impacted and contributed to my study!

Orhan Sivrikaya

May 2013, Okan University, İstanbul

CHAPTER 1 INTRODUCTION

1.1 Objective

Accuracy in estimating airline market size is a key element while an airline is preparing its short term or long term business plan whether it is an incumbent or a startup company. Fleet planning, aircraft scheduling and routing, design or fine tuning of a flight network and human resource planning have at their heart a reliance on an accurate forecast of demand.

Main drivers of airline traffic generation are classified by two groups which are geo-economics and service related factors. Although the purpose of air travel can vary in a wide scale depending on passenger profile, selection of a trip destination is expected to be explained with such determining factors. The model is based on air travel demand forecast for individual origin and destination market of Turkey. Any city pair for which direct or connecting flights are available are treated as an individual market.

The objective of this study is to examine the demand forecast for domestic air transportation in Turkey, by developing predictive models based on the method of multiple regression analysis against artificial neural network approach. The estimated model comes into prominence since it does not only predict the pattern for actual traffic results but it also estimates the demand volume for each off-line city pair without past data. Therefore, the model can be used for micro level evaluation studies for various purposes such as feasibility study for a new airport construction or launching a new route in flight network etc.

Time series analysis, gravity model, grey theory and artificial neural network are the tools commonly used for air traffic demand forecasting. However, time series and grey models are suitable for trend prediction and not useful to include non-existing sample variables in the model. Therefore, in this study, multivariate semi-logarithmic regression as a traditional forecasting technique versus artificial neural network as a novel forecasting technique were applied by identifying, analyzing and interpreting the factors which influence air passenger demand. Different aspects of the two models such as accuracy level and their inferences were compared.

1.2 Problem Statement

Accomplishing a good fit in prediction has two important aspects; 1- Identification of determining factors 2- Establishing an accurate estimation model. New software tool using stepwise regression technique is able to choose the right combination of variables out of candidate variables. However, the approach to find out potentially significant variables out of the candidate list is the main concern. Purposes of trip can be categorized by four subgroups which are tourism, business, VFR and education. It is not relevant to classify the O&D's by those subgroups because they are all present for each city pair with some percentages of composition. Therefore, the explanatory variables are to be found out between socio-economic or industry related parameters so that they would be expected to represent those basic purposes of travels. For example, majority of traffic for the city pair İstanbul and Bodrum is assumed to be holiday purpose. It requires to drill down some characteristics of the concerning cities for the most representative explanatory variables so that the estimation model will distinguish them resulting in accurate prediction through the holiday routes.

Actual traffic and potential traffic are indeed two different concepts. In this study, air traffic demand refers to potential traffic as much as the models predict. Actual traffic is considered as realized portion of potential traffic. Hence, actual traffic is obviously supposed to be smaller than the potential traffic. For example, if there is a shortage in capacity as a result of any facts such as monopoly, high yield policy or lack of aircraft availability etc., demand would be spilled in most of the time. In this case, actual traffic would be much less than potential traffic. Furthermore, it does not mean that air traffic demand is insignificant for the city pair with no air service. That is why time series analysis is not applied in this study since it is based on actual results.

Although a multi linear regression model may be found a good fit for estimation, prediction performance can be deteriorated when different range of the concerning explanatory variable starts to be observed. Like a hidden or imperceptible heart attack which would most likely be a serious health problem in future. The stable variable which actually has a nonlinear behavior on the model may not be noticed with a small range of changes in the linear model version. Alternatively, the variable may have threshold values on the scale to influence the model in different forms. In the light of aforementioned points of view, multivariate semi-logarithmic regression model was first established to develop the framework of the estimation models in comparison with artificial neural network, which is able to overcome with possible nonlinearity aspect of the model, to seek progress in prediction performance.

1.3 Limitation of Airline Demand Forecasting

Aviation demand models can be very useful forecasting tools, but it is important to be aware of their limitations. First, all models does not necessarily provide a complete solution. They intend to transform a large and complex system to a relatively few mathematical equations that describe the most important interactions. Many factors must be eliminated, either because they are impossible to formulate mathematically or because having them in would make the model ambiguity and too complicated to use. There are other factors kept out not by design but by inadvertence because, with the current state of knowledge about the relationship between aviation and underlying economic and social indicators, we are simply unaware of all the determining factors that drive demand for air transportation. To overcome with such structural weakness, the model builder calibrates and tests the model by using data from past years to check if the historical record can be accurately reproduced. If the model is well established and its mathematical relationships are a good representation of reality, using data from some past year in the equations will provide a forecast of the aviation activity that actually existed. Such testing enables the forecaster to ensure that for some future year the model will repetitively predict aviation activity if the correct values of input variables are used. Nevertheless, the correct values of explanatory variables such as GDP or population for any future year are themselves unknown, and uncertainty is simply reflected from the behavior of the economy to behavior of the airline industry at large.

Models usually assume that the future will be very much like the past. If the real world situation changes substantially, the model would be respectively less accurate. Such changes might include unexpected economic shocks, such as the fuel crisis of 2008, or longer term restructuring of the market or the economy. For example, aviation models were generated to predict the behavior of a regulated industry with stable fares and routes tend to be less accurate now due to intensive liberalization in aviation industry, yielding price competition and free entry into new markets. Furthermore, because models are, at best, only partial representations of the reality, it would be difficult to predict which changes in air travel behavior or how much economic conditions will be important in the future or how they should be incorporated in a model. A major problem among forecasters is distinguishing among relationships that will persist and those that will not.

From this, it is obvious that aviation demand models highly depend on underlying assumptions about socio-economic and airline industry and future conditions. The model itself may precisely show relationships between air transportation and the state of the economy or the structure of the aviation industry, but if the assumed states of these variables at some future time are too optimistic, too pessimistic, or simply inconsistent with the course of events, the resulting forecasts would be useless. It is reasonable to conclude that, even with the current limitations of the model builder's art, the inaccuracies due to the structure of forecasting models are generally smaller than those induced by erroneous input assumptions. An aviation demand model is no more robust than the assumptions on which it relies. Therefore, assumptions are the most fragile and sensitive part of demand forecasting process.

The limitations, biases, and characteristics of a forecast would depend as much on who is doing the forecast as on the particular method being used. Airport authorities, manufacturing companies, government agencies, airline companies, and industry associations all make forecasts to enable them to plan for the future and to help them to revise plans and programs. There is, in many cases, a natural tendency to mislead on the side of optimism in order to serve for the future interests of the agency producing the forecast. For example, local airport authorities planning an expansion project may tend toward an unreasonably high appraisal of the overall growth prospects in the local economy and, therefore, future passenger or cargo traffic. As a result, basing decisions to construct or expand facilities on these forecasts may come up with excess capacity or premature investment of capital. Nevertheless, this may be less harmful than relying on a forecast which is too low. Slight overcapacity and overestimation of demand is considered by most airport planners as preferable to congestion, delay, and perhaps deterioration of service and safety.

CHAPTER 2 AIR TRAVEL DEMAND MODEL

Big changes in airline passenger traffic in Turkey create a challenge to testify robustness of any claiming estimation model. Airline traffic is most of the time considered as a significant indicator for the performance of the nation's entire industry since it is highly correlated with the number of business activities. However, macroeconomic or demographic changes do not seem to be responsible for entire boost in the air travel demand since airline industry has demonstrated much higher growth rate than macro-economic indicators in Turkey. Competition doubled or tripled available seat capacity on some routes; therefore it required a different strategy for competing airlines to generate additional demand to achieve in satisfactory load factor which is a key performance indicator for airline profitability.

Airline deregulation has changed the national air service network from a stable system of routes served by established carriers to a fluid market- place where carriers frequently adjust flight destinations, level of service, and fares. The older airlines have abandoned some markets and begun service to others. New entrants have taken over some of these abandoned routes and established themselves in the dense markets where they see a competitive opportunity.

The course of the industry since deregulation has been one of uneven expansion and sometimes contraction. The principal effect on airports is that sudden and less predictable changes have taken place in the air carriers serving the airport, the level of service provided, and the facilities needed to accommodate them. Some cities have experienced a general improvement in air service since deregulation, but not all have benefited from an unregulated environment, and some have suffered almost complete loss of air service.

2.1 General Demand Forecasting

Air travel demand forecast is, in essence, a carefully formed prediction about future air traffic. Its primary use is in foreseeing future needs and generating inputs for decision support system. Any of several alternative methods may be used, with results that will vary widely in terms of scope, time scale, structure, and detail, but all have certain common type of features. Mainly, forecasts are based on assumptions about the relationship of the past and the future in that they postulate that certain measurable historical events or conditions. Analysis of these historical factors, usually by some sort of mathematical formulation of data, allows the forecaster to express expectations in terms of some measure or index of aviation activity.

In establishing an estimation model to prepare a forecast, the forecaster has at his disposal two basic types of input data. He may choose data on aviation activity itself and use historical data trends to project future activity. In principle, this approach assumes that the best predictor of future aviation demand is past aviation demand. Alternatively, the forecaster may choose data related to underlying economic, industry related, social, and technological factors that are presumed to influence aviation demand, categorizing them as independent variables that can be used to predict demand as a dependent variable.

The outputs of the forecast are measures of aviation activity such as aircraft movements, number of revenue passengers, freight ton-miles, and number of aircraft in the fleet, number of aircraft operations or aircraft types in fleet composition. The range and scope of forecasts can vary greatly, depending on the purpose they are to serve. The geographical scope may be international, domestic, regional, or specific to a particular market or airport.

The forecasting horizon may range from a few months up to 20 years, again depending on the objective of the forecast. Airlines, for example, keep using short-term projections of passenger demand in order to allocate optimum capacity for reservation booking classes on daily basis. Airport planners, on the other hand, use very long-range forecasts, up to a period of 20 years, as a basis for major decisions relating to land acquisition and airport development. Between these extremes, forecasting horizons of 5 or 10 years are common for planning changes and improvements of airport facilities.

There are two different basic approaches to aviation demand forecasting:" top-down" or "bottom-up." The top-down approach begins with the largest aggregates of economic and statistical data and seeks to provide a general picture of aviation demand including the country or the entire system of air travel routes and facilities. Once the aggregate forecast has been accomplished, portions of the total volume of traffic can be disaggregated to specific industry segments or geographical regions based on historical shares or assumed growth rates.

The bottom-up approach, in the contrary, begins with data for a specific regional area and develops a forecast of aviation demand at a particular airport or city pairs. Where good data are available and the economy of the region is developing in an orderly way, this approach can accurately approximate the reality of the area under study. In some cases, a number of such bottom-up forecasts may be combined to make an aggregate forecast for a larger area.

Whether “top-down” or “bottom-up” aviation demand forecasting as practiced today utilizes a wide variety of methods. The characteristics, limitations, and typical applications of these quantitative methods are described below.

2.1.1 Time Series

A commonly used forecasting method is the extrapolation from the past, where the forecaster assumes that major trends, such as traffic growth or market share, will continue consistently so that the future will be similar to the past. Historical data for some certain period are collected and examined to detect a trend line, which is then extended to some point in the future, using either sophisticated mathematical procedures or simple estimation of the most likely course. This method is often used for short-term projections (1 or 2 years) where basic conditions are unlikely to change much. For example, airline yield management system is usually based on time series forecasting which anticipates the number of passengers on board by weighing between the historic and advanced booking data. However, a basic handicap of time series extrapolation is that it does not take into account underlying economic, social, and technological factors that affect aviation and which are subject to change.

2.1.2 Market Surveys

This method has been used extensively in the past when data availability was almost impossible. Forecasters used in-flight passenger surveys to collect information on point of origin, choice of airport in the metropolitan area, choice of ground access mode, ground access travel time, destination, purpose of trip, and other factors that can be employed to predict air travel behavior and induced demands on aviation facilities. These data are grouped in a travel market model made up of miscellaneous socioeconomic “cells” defined by age, occupation, income, and trip purpose. The growth rate for each cell is forecasted by straightforward econometric techniques.

The market survey method, while it provides a highly detailed forecast of travel, has some significant handicaps. Data collection is complicated, time-consuming, and expensive. Since the sample is collected in a relatively brief period, it may not be truly representative of long-term travel patterns and preferences. Airlines, which serve as collectors of the data, are reluctant for confidentiality reasons to relinquish control of survey results which they consider their proprietary and commercial secret.

2.1.3 Ratios

Some aviation authorities, government agencies and industry groups make forecasts by the relatively simple approximation of assuming a ratio between national “top- down” traffic forecasts and their own segment of traffic. This method is generally used by airports that lack the funds or expertise to make independent econometric forecasts.

2.1.4 Scenarios

The scenario method is often used to show the variation due to differing assumptions about future conditions, thus specifying the-range of uncertainty. The values of input variables in an econometric model, for example, are in themselves simply guesses about the future behavior of the economy. Rather than depend on a single “best” estimate of GDP in coming years, the forecaster may elect to construct several scenarios to predict the behavior of the aviation industry under a range of likely economic conditions.

One of the drawbacks of the scenario method is that the range between high and low estimates can be so large that the forecast loses practical value as a guide to planning.

2.1.5 Econometric Model

The econometric model is by far the most commonly used method for forecasting aviation demand. It is a mathematical representation of air traffic or its constituent parts and those independent variables of the national economy which are thought to influence traffic growth. Econometrics is the statistical technique used to quantify these relationships. The mathematical equations of the model relate economic factors or industry related drivers to the level of aviation activity, based on observation of past behavior of both the economy and the aviation industry. The regression model may also be established so as to reflect the effects of airline service specific factors such as average ticket fare, the number of airline choices, etc. Airlines used to employ econometric model for long term fleet planning. Passenger traffic growth for each flight route is estimated for the forthcoming years then the right number of aircrafts and the best aircraft type are determined by fitting estimated air passenger demand with the minimum cost of aircraft operations on the concerned flight routes.

2.1.6 Gravity Model

The gravity model was first developed in the sociological and marketing fields to estimate various forms of human interaction. The technique was later adapted by traffic engineers to estimate travel behavior. It is predicated on the assumption that travel behavior obeys a law analogous to the law of gravity, in that attraction between cities varies directly with population and inversely with distance. Thus, two large cities located near one another have a strong mutual attraction and form a very dense transportation market; small cities located far apart have little travel between them. The gravity model uses socioeconomic data for each pair of metropolitan areas to predict the level of transportation activity between them. The equations often contain terms to describe the special attractiveness of each city for different types of personal and business trips.

The gravity model can be easily adapted to multivariate linear regression by logarithmic transformation of corresponding variables.

2.1.7 Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing analogy that is inspired by the way biological nervous systems, such as the brain, process information. It consists of a large number of highly interconnected processing elements (neurons) working in harmony to solve specific problems. ANNs, like people, learn by example.

Air transportation system is very difficult to model with traditional approach, either because the interactions among the different system components are not fully interpreted or because one is dealing with a lot of uncertainties. For such complicated systems, building empirical models, based on observed data are, may be the only option remaining. NNs, given their universal function approximation capabilities, are the best alternative tools for building such models.

Neural networks have broad applicability to real world business cases. In fact, they have already been successfully utilized in many industries. Since neural networks are best at identifying patterns or trends in data, they are well suited for prediction or forecasting needs including:

1. sales forecasting
2. pattern recognition
3. customer research
4. data validation
5. risk management
6. target marketing

2.2 Air Transportation Progress

Aviation represents the only transportation network across the globe and it is crucial for global business development and tourism enrichment. Air transportation is one of the most important services to provide both significant social and economic benefits. By serving tourism and trade, it simultaneously contributes to economic growth. It also provides jobs and increases government tax revenues. Air transportation is inevitably essential for the fast movement of people and cargo shipments around the world. Finally, air transportation improves the quality of people's lives by broadening their leisure and cultural amenities. It gives a wide choice of holiday destinations around the world and is a possible way to visit distant friends and relatives.

The use of commercial aviation has grown significantly over the last few decades since the first jet airliner flew in 1949. This rapid growth is based on a number of different factors. First, rising disposable income and quality of life in many parts of the world have encouraged more people in these areas to travel and explore overseas opportunities. Second, deregulation of aviation laws, and bilateral and open-sky agreements between governments have opened new markets for airlines, which make travel easier and affordable. Third, demand is increasing because of growing confidence in aviation as a safer mode of travel. Fourth, increased efficiency and increasing competition have reduced world airfares and the cost of travel. Last but not least, globalization has increased the average distance traveled, as people do business in countries which now have improved political and social environments. The impacts of these factors are expected to continue to be effective, however, at different levels in different parts of the world. The number of air travelers and the volume of air cargo are expected to continue to grow, increasing the pressure on all the contributors to the air transportation service to expedite business opportunities around the world and efficiently manage their service.

The recent liberalization of EU Skies has caused the disappearance of “flag carrier” status in relation to intra-EU services and state-owned carriers are now required to compete on equal terms with privately owned carriers. Given that a high number of airlines operating within the EU are either entirely or partially owned by the state, state aid is an issue of crucial importance: more so in the context of a liberalized market.

The liberalization of air travel is very important not only for aviation industry itself but also for nationwide economy as a whole. The airline industry plays a very important role in the whole transport sector and its contribution is also getting more and more important as the speed and efficiency become basic need, as well as the lower prices after the liberalization of the air travel market are making the industry more attractive for consumers and businesses. The liberalization of air travel industry has not only brought competition and lower fares but also a significant increase in bilateral relations between cooperating countries worldwide.

The importance of liberalization and competition has long been similarly recognized in Turkey as well, and in fact the existing law was not enforcing a monopolistic market, however aviation policy had made free entry almost impossible. The privatization efforts of Turkish Airlines had also shadowed the progress of liberalization since priority was given to the former.

The liberalization of Turkish domestic passenger flights market has been eventually realized by a policy change of the Ministry of Transportation in October 2003. Up to that date, there were no legal restrictions for the private airline companies to provide scheduled services on domestic routes. However, the state aviation policy was based on the protection of Turkish Airlines from competitive pressure. Therefore, the General Directorate of Civil Aviation abstained from granting operation permit for the private airline companies on domestic scheduled routes.

As a result of a policy change initiated by the government, parallel to the privatization efforts of Turkish Airlines, the domestic market has been opened to competition in October 2003. Before this policy change, the private enterprises could provide services (charter or scheduled) on domestic routes where Turkish Airlines did not have flight or could not provide sufficient service. Since these routes were not lucrative enough, the domestic passenger did not demand private airline company's services as expected; therefore their domestic services were limited to a number of seasonal touristic flights.

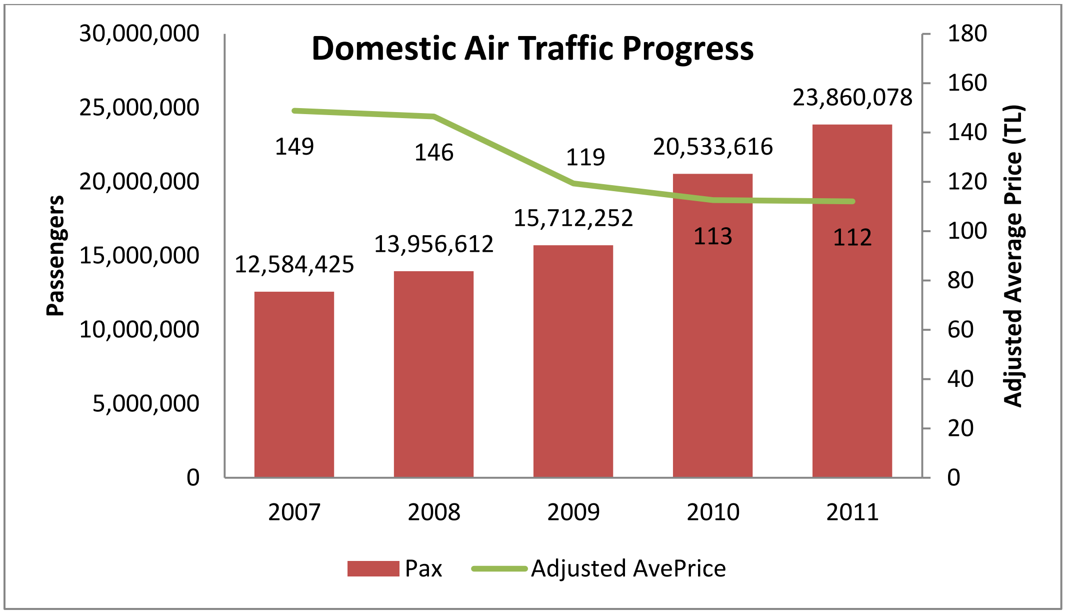
As the liberalization was taken into practice, a number of existing private airline companies having international flights (both charter and scheduled) has begun to provide services on domestic routes as well, and a few startup companies has been established. As a result of the liberalization, significant progress in number of planes and seat capacity was realized by the private airline companies carrying passengers on domestic and international routes.

Turkey is spread over a wide geographical area, and both railways and roadways are not adequately constructed for all directions between the regions. Therefore, air transportation is supposed to have more shares out of total statistics of domestic transportation modes covering all possible city pairs. This fact contributed air travel demand to demonstrate such a remarkable jump when the relative price became affordable amongst all domestic transportation modes of Turkey.

Figure 2.1 shows that domestic air traffic has increased by a compound annual growth rate of 17.3 % between 2007 and 2011. While the number of air passengers was gradually growing through the years, average ticket price which was adjusted by inflation rate continuously decreased. This fact implies an inverse relation between air travel demand and average ticket price as it is evident that low cost carriers succeeded to enlarge the market size by low fare policy in recent years.

The deregulation of air transportation market in Turkey in 2003 has started revolutionary changes in the airline industry. New government having the target to increase the portion of air travel out of all modes of local transportation has encouraged more airline companies to enter into the domestic market. New policy eventually enabled them to offer more attractive ticket prices due to tax cutting measures specific to the airline industry. Liberalization has consecutively lowered prices significantly in the domestic passenger flights market.

Figure 2.1 Domestic Air Traffic Trend



Price oriented competition has worked very well, noticeably increasing airline passenger traffic. Low cost carriers have contributed to a continuous two-digit growth by stealing passenger traffic from bus transportation companies. Low fare policy was successful to generate new passengers. Low Cost Carriers such as Pegasus, Sun Express, Atlas jet and Onur Air played important role in dramatic increase in market size. While the price gap between bus and air travel is shortening, significant amount of bus passengers start to use air transport mode resulting in continuous growth of domestic air travel with accelerating rate. Turkish Airlines as a legacy carrier has performed structural changes at home market by applying new tactic and strategies such as dynamic pricing policy and sustainable growth strategy to make economies of scale working. Although Turkish Airlines has not changed its business model and stayed as a full-fledged carrier, it achieved significant amount of cost saving as a result of continuous progress in productivity.

Big changes in airline passenger traffic in Turkey created a challenge in applying an accurate demand forecasting model built to estimate the number of air passengers. Macroeconomic or demographic changes do not seem to be the only responsible drivers for such amount of increase in air travel demand. Competition resulted in the increase in available seat capacity with new comers on some routes. This, in turn, forced airline companies to formulate a different strategy to generate additional demand to utilize the extra capacity in order to achieve in satisfactory load factor, a key performance indicator for airline profitability.

2.3 Literature Overview

It has been seen throughout the results of the previous research in the literature that one of the most important issues to develop a predictive model is to choose the right combination of the variables which represent the determining factors involved in the model. These variables are categorized by two subgroups (Carson et al., 2011): 1.Geo-economics Factors: which consist of geographical characteristics, economical activities, social factor etc. 2. Service Related Factors: which are related to airline dependent factors.

The other prominent aspect of model generation is the level of forecast which can be classified by two groups: 1. Microscopic Model: Airport specific or city pair specific data are involved such as the total number of incoming and outgoing passengers per particular airport or per city pair. 2. Macroscopic Model: Region or country specific data are involved such as aggregate number of passengers in a region or country regardless of origin or destination city.

Local area information appears to be more relevant in determining local O&D travel than of national information such as gross domestic product (Bhadra, 2003). Factors representing the industry associated with factors that are specific to a region are critically important for the existence of aviation companies as a whole. All these imply that local economics play a significant role in the development of air transport industry.

Due to lack of data, aviation companies used to apply to top-down macro econometric models in demand forecasting. These models are unable to describe and analyze complex and dynamic route networking (Bhadra, 2003). Recently, structural econometric models at micro scale such as O and D (Origin and Destination) level can be used as a result of significant advances in Management Information System (MIS) of the companies.

A variety of different models exist for airline passenger volume estimation. Since no single model guarantees accuracy, airlines compare forecasts from several different models. Amongst the existing set of forecasting methods, the most commonly used demand models are of the simple gravity type formulation (S.C. Wirasinghe et al., 1998). Gravity type of functional formulation can be easily transformed to logarithmic expression so that multivariate regression model is applicable.

A comprehensive review of researches on air passenger demand forecasting was presented by Jorge Calderon (1997). A demand model for scheduled air carrier services was formed for the whole network of international European routes in 1989. The estimated model covered the variables describing both the geo-economic characteristics of the region where transportation took place and the patterns of airline service, as determined by flight frequency, aircraft size and air fares. Brons et al. (2002) collected 37 studies and 204 observations, and showed that the demand for air travel is largely determined by the spending capacity of customers. Hsu and Wen (1998) used grey theory to develop time series GM (1, 1) models for forecasting total passenger and 10 country-pair passenger traffic flows in Asia-Pacific market. A series of models capable of forecasting airline city-pair passenger traffic, designing a flight network of airline routes and determining flight frequencies on individual routes were developed by applying grey theory and multi-objective programming (Hsu and Wen, 2000). Local area information appears to be more relevant in determining local O&D travel than of national information such as gross domestic product (Bhadra, 2003).

Later, Grosche et al. (2007) introduced two gravity models for the estimation of air passenger volume between city-pairs. The models consisted of variables describing the macro economic activity and geographical characteristics of city-pairs instead of variables describing air service characteristics. Booking data of flights between Germany and 28 European countries were used for calibration. The model moderately showed a good fit to the observed data which contains 1228 city-pairs with 137 embracing cities.

Weatherford et al. (2002) presented the first artificial neural network (ANN) forecasting model as applied to the airline industry. It compared the new method with the traditional forecasting techniques. The research indicated that the neural network structure is able to perform complex forecasts much more easily than any traditionally used forecasting technique. Recently, Kuo et al. (2010) employed artificial neural network to establish mathematical model with multiple inputs and multiple outputs. The results show that the novel model may accurately forecast the air travel demand in routes network.

2.4 Components of Air Transport Demand

The determining factors with respect to geographic scale; local, national, and international; sector of industry; and different components of air transport will all act to varying degrees of impact and play a complementary role to estimating the air transport demand.

On the local and regional levels, the socioeconomic and demographic variables and the shifts and directions the economy takes would take the major part in estimating the number of passengers within the region or airport.

On the national level, the state of the national economy and the state of the airline industry are the major factors that would influence aviation demand. Other factors include geographic and demographic distribution of demand, technological improvement in the industry, and probably government policy on sensitive environmental issues.

Internationally, bilateral agreements, state of global and regional economies, political considerations, regional turmoil and internal security, airline globalization, cultural and social ties between nations, and advances in aeronautics, telecommunication, navigation, and surveillance technologies all may dictate the size and kind of international air travel.

In terms of components of air travel, the airports and airlines distinguish between the origin and destination (point to point) passengers and the connecting passengers. The split of passenger demand between these two basic types of demand may affect how the airlines operate, but most importantly how the individual airport facilities are structured.

O and D (Origin and Destination) passenger demand is the passenger trips originating or terminating at an airport. It indicates the passenger demand directly associated with the airport/region local socioeconomic and "propensity to travel" characteristics.

Since deregulation of the airline industry in the late 1970s, airlines have redesigned its flight network with hub and spoke system as a strategy to gain market share and operate more profitably. Connecting passenger demand means any passengers on flights to/from the passengers' origins and destinations (or city pairs) who have to go via a third airport depending on the airline flown the airline hub. Establishing a clear understanding of airline hub planning considerations and quantifying the connecting element of passenger demand have always been problematic and difficult to make a model. It is known that establishing hubs has always been a closely guarded decision by airlines managers. It is a decision predominantly related to the airline's business model and marketing strategy. But in broad terms, the conditions that favor an airport to be established as a hub by an airline may include geography and orientation of the network structure, airport infrastructure capacity, strong OD base, network expansion, aircraft fleet of the airline, competition with other airline and hubs in close proximity, the airline adding opportunities to improve its profitability and market dominance, and potential for establishing an international hub.

An important aspect of airline flight network is the cooperation possibilities with commercial air carriers and regional/commuter airlines on well-integrated route, networking to achieve in good penetration to communities and effective market coverage. In the past, a variety of strategies were used by airlines to gain more route structure integration, yielding to market control and enter new markets. Strategies such as code sharing, revenue management systems and acquisition of regional airlines helped, providing better connections and service frequencies to passengers and higher load factors and hence revenue maximization to the airlines.

In addition to airline demand forecast, airport demand forecast may be also influenced by intermodal interactions with other transport modes that may be competitive with or supplementary to air travel. Therefore, in conducting air travel demand forecast for a particular airport or regional route, this aspect has to be carefully taken into consideration. Implementation of competing or supplementing ground transport modes will largely depend on such evaluation as the geographic region's ground transport network, technological capabilities in ground modes, financial and economic feasibility of all travel modes, future availability of options and their costs of energy, perception and acceptability by the traveling public, and the environmental and social impacts of the modes.

2.5 Factors Contributing to Air Transport Demand

Demand for air travel industry is invariably affected by a variety of causal variables. In order to establish a reliable and accurate estimation model, these variables should be unambiguous and measurable and the available data should reasonably conform to mathematical formulation and statistical analysis. These causal variables are intrinsic to models that enable future estimates of demand. They reflect the various components of air transport demand represented in the respective demand models. Causal variables typically used for demand forecasts, their influence on demand, and corresponding model type are shown in Table 2.1.

Air passenger demand is highly correlated to a region's population and the motivation of individuals to travel, in other words their propensity to travel, as well as socioeconomic activities and measures that support travel and the availability of related services and infrastructure. The underlying assumption in all forecasts is the strong correlation between demand and trip-generating factors that are derived from historical data and this correlation should be consistent through the forecasting horizon. Expected future demand environments expressed as forecasts of such factors as ticket price, level of service quality, gross domestic product, and so on, are all inputs to the forecasting process. An underlying assumption in all forecasts is that forecast models hold in the future as long as associated assumptions related to all factors hold in the future as they do at past and present. For examples, econometric forecast models are built by using time series historic database or industry cross- sectional data. Both availability and accuracy of the data used are critically important to this process both for airlines and for airports.

In conducting forecasts of air travel demand, the following factors are considered:

1. Availability of capacity; airports and airspace limitations
2. Macroeconomics situation; locally, nationally, and internationally
3. Socioeconomic and demographic variables of the airport region
4. Economic factors directly related to airlines operation cost at the airport
5. Competition between airlines serving the airport as well as competition between the air and other modes of transport
6. Environmental and political constraints on the air transport system and airline industry

7. Technological progress in aeronautics, telecommunication, air navigation, and other related fields

8. Overall safety, security, and convenience of air travel

Table 2.1 Demand Variables and Application

Type of influence	**Variable**	**Application**
Market size and consumer purchasing power	Population or number of households Gross domestic product Exports Imports	Passenger forecasts All types of forecasts Outbound international freight Inbound international freight
Ethnic (or linguistic) ties between areas	Proportion of population of intercity immigrants	Passenger forecasts for route or group of routes
Price of air service	Published tariffs Revenue yield	Route forecasts All types of forecasts
Quality of air service	Departure frequency Number of stops or connections on a route Travel time	Scheduled route forecasts Scheduled route forecasts Route Forecast
Access to air transport services	Number of destinations served Proportion of market within a certain distance or travel time from airport	Regional forecasts Airport or route forecasts

The forecaster must pay good attention to the manner in which air travel forecasts are presented. Reasonable presentation of forecast is vital to acceptance of the forecasts and success of the estimation model. Generating the forecast model, including performing the required statistical tests, may not be enough for acceptance.

The demand forecaster must also consider the following aspects:

- Statement of purpose for the forecast
- Relation of the forecast being presented to the entire region forecasting process
- Description of the air travel environment and the unique market situation
- Forecast methodology, including approach, use of assumptions, model mechanics, and reasons to adopt the particular approach
- Assumptions related to determining factors affecting demand and justifications for the assumptions
- Historical records and databases for the causal variables making the model, quality of these data, and data sources and definitions
- Discussion of the accuracy of the forecast model indicating the range of uncertainty and model boundaries

From a quantitative perspective, the demand forecast is not only a final objective on its own, but also a part of a larger goal and greater purpose. The primary use of forecast is to guide the airport planner or airline scheduler to develop a strategy for tackling several important interrelated tasks. These tasks include developing new or expanded facilities, determination of financial feasibility, mitigation of environmental impacts, and conducting a complete master plan for the airport or fleet development.

Specifically, demand forecast is central to the formal airport master plan, which constitutes demand versus capacity analysis, sizing facility requirements, airport development conceptual plans, and economic-financial feasibility in addition to airline fleet plan.

In general, demand forecasting is the backbone to the plan of future development for the airport infrastructure or airline fleet size and composition.

2.6 Demand Model

The novel artificial neural network was developed against, the traditional forecasting model, multivariate semi-logarithmic regression in order to compare prediction accuracy for the number of passengers between domestic city pairs in Turkey.

The statistical tool SAS Jmp 10 is utilized in this research. Semi-logarithmic regression model, which is derivative version of gravity model, is generated by using sample data of 2011 and prediction performance of the generated model is assessed by using testing data of 2010.

2.6.1 Gravity Model

A comprehensive review of researches on air passenger demand forecasting was presented by Jorge Calderon (1997). A demand model for scheduled air carrier services was formed for the whole network of international European routes in 1989. The estimated model covered the variables describing both the geo-economic characteristics of the region where transportation took place and the patterns of airline service, as determined by flight frequency, aircraft size and air fares. Brons et al. (2002) collected 37 studies and 204 observations, and showed that the demand for air travel is largely determined by the spending capacity of customers. Later, Grosche et al. (2007) presented two gravity models for air passenger volume forecasting between city-pairs. The models used mainly geo-economics variables rather than service related factors. The model moderately showed a good fit to the observed data which contains 1228 city-pairs with 137 embracing cities.

The gravity model takes the form:

…

This model assumes that the marginal effects of each variable on demand are not constant but depend on both the value of the variable and the values of all other variables in the demand function (Aderamo 2010). In other words, the explanatory variables affect demand in multiplicative manner. Partial derivation of any independent variable proves aforementioned relationship. However, this model can be made suitable for multiple regressions by applying logarithmic transformation. Logarithmic form of the gravity model takes the form:

, where

It is obvious that interdependency is resolved in this form so that multiple regression model can be applied.

Following the demand function used by Aderamo (2010), this study specifies the demand for air travel service as a function of geo-economics and industry related factors depicted in Table 3. Based on the theoretical foundation, the original form of the demand model in this study is written as follows from origin i to destination j:

.

The linearized form is then specified as:

,

Where i and j represent origin and destination city, and are the coefficients of independent variables. The variables are defined as follows for each i and j:

: Logarithmic transformation of the number of air passengers,

: Matching probability of travel plan with available weekly flights,

: Elapsed time between the start and end point of the journey,

: Dummy variable indicating connection type; 0 for direct service and 1 otherwise,

: Number of airlines serving on the route,

: Logarithmic transformation of the average ticket price,

: Continuity of flight service in number of months in a year,

: Logarithmic transformation of the product of the numbers of available beds in tourism facility,

: Logarithmic transformation of the product of cities population and

: Road distance.

The statistical tool SAS Jmp 10 is utilized in this research. Semi-logarithmic regression model, which is derived version of gravity model, is generated by using sample data of 2011 and prediction performance of the generated model is assessed by using testing data of 2010.

The service specific factors, which are globally considered as airline service related aspects representing air travel demand attractiveness, travel match, travel time, transit, price, airline count and schedule consistency are embedded in the model in addition to the macro economic factors which are population, distance and the bedding capacity of the cities.

2.6.2 Artificial Neural Network

One of the recent technological advancements has begun to provide an airline with an array of opportunities in pattern recognition is the neural network (Weatherford et al., 2002). There is literally numerous number of neural network structures that can be designed. The structures can vary in the number of layers, the number of neurons, the number of inputs and outputs, type of activation functions, feedback information, etc. One of the most commonly used supervised ANN model is back-propagation network (BPN). BPN is a layered feed-forward supervised network which is suitable to establish air passenger forecasting models comparing with gravity model in this study.

The neural network is a technology that learns like human brain: it learns from being trained. Training occurs when patterns of given inputs and known outputs are repeatedly applied to the network. Through the repetition, the network iteratively adjusts each weight until the difference between the desired or expected output and actual output is below a predetermined value. The difference between the desired and actual output is called the error (Proctor, 1992).

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyze. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

Other aspects include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.

2. Self-Organization: An ANN can create its own organization or representation of the information it receives during learning time.

3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

4. Fault Tolerance via Redundant Information Coding: Partial destruction of a network leads to the corresponding degradation of performance. However, some network capabilities may be retained even with major network damage.

There are two types of ANN architectures:

1. Feed-forward networks:

Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.

2. Feed-back networks:

Feedback networks can have signals travelling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent, although the latter term is often used to denote feedback connections in single-layer organizations.

Figure 2.2 Artificial Neural Network

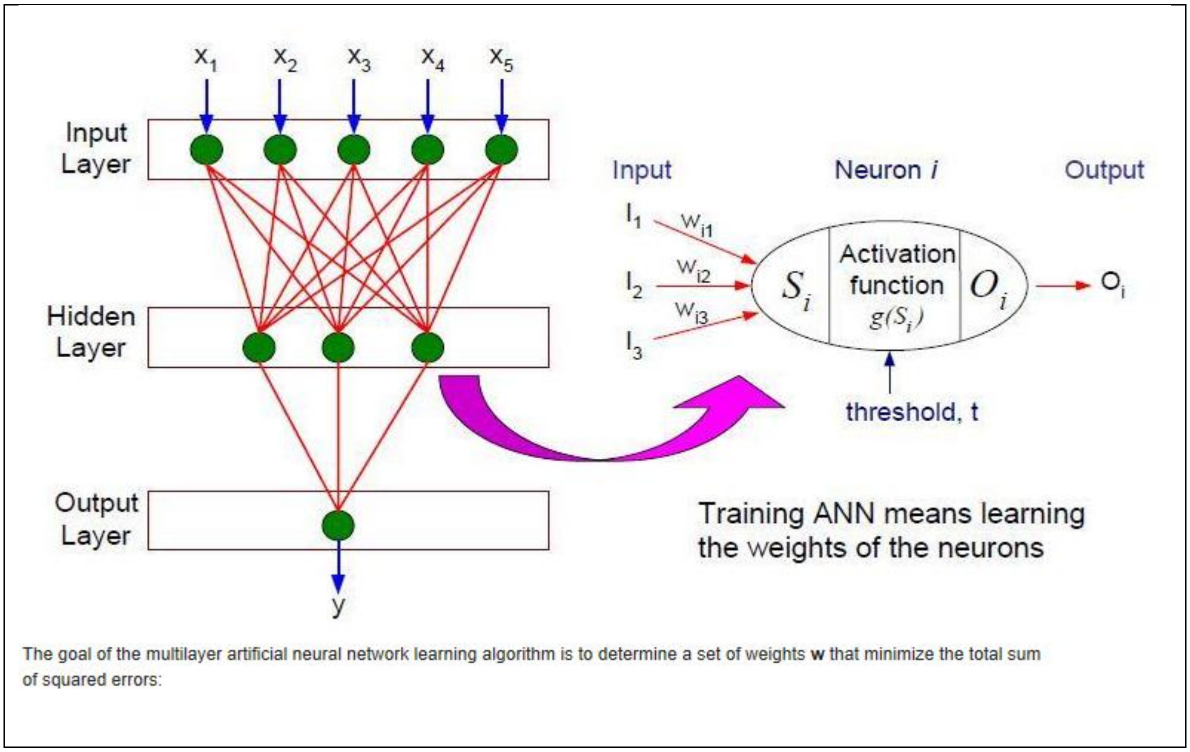


Figure 2.2 shows the commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units.

The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

This simple type of network is interesting because the hidden units are free to construct their own representations of the input. The weights between the input and hidden units determine when each hidden unit is active, and so by modifying these weights, a hidden unit can choose what it represents.

We also distinguish single-layer and multi-layer architectures. The single-layer organization, in which all units are connected to one another, constitutes the most general case and is of more potential computational power than hierarchically structured multi-layer organizations. In multi-layer networks, units are often numbered by layer, instead of following a global numbering.

All learning methods used for adaptive neural networks can be classified into two major categories:

1. **Supervised learning :**

It incorporates an external teacher, so that each output unit is told what its desired response to input signals ought to be. During the learning process global information may be required. Paradigms of supervised learning include error-correction learning, reinforcement learning and stochastic learning.

Important issue concerning supervised learning is the problem of error convergence, i.e., the minimization of error between the desired and computed unit values. The aim is to determine a set of weights which minimizes the error. One well-known method, which is common to many learning paradigms, is the least mean square (LMS) convergence.

2. **Unsupervised learning:**

It uses no external teacher and is based upon only local information. It is also referred to as self-organization, in the sense that it self-organizes data presented to the network and detects their emergent collective properties.

The behavior of an ANN (Artificial Neural Network) depends on both the weights and the input-output function (transfer function) that is specified for the units. This function typically falls into one of three categories:

1. linear (or ramp)
2. threshold
3. sigmoid

For linear units, the output activity is proportional to the total weighted output. For threshold units, the output are set at one of two levels, depending on whether the total input is greater than or less than some threshold value. For sigmoid units, the output varies continuously but not linearly as the input changes. Sigmoid units bear a greater resemblance to real neurons than do linear or threshold units, but all three must be considered rough approximations.

The training of the neural network assigns more positive weights to more important data (excitatory) and assign more negative weights for less important data (inhibitory) (Hruschka,1993). Every neuron sums the products of the inputs (X1, X2, X3…) and associated weights (W).The sum of each neuron is then compared with its threshold value to determine whether that neuron will be activated. Neural networks are identified by the type of the neuron’s activation function with its classification of characteristics and abilities.

CHAPTER 3 EMPIRICAL ANALYSIS

This study aims to find out the determining factors turning potential demand into air travel passengers. Population, gross domestic product per capita and distance are considered as the leading geo-economics dynamics behind air travel demand as depicted in Table 3.1. Average price has a stimulating effect on airline demand as Brons et al. (2002) pointed out that ticket price is an elastic driver for airline demand generation. However, most of the related studies used top-down macro econometric models. Therefore, it may be claimed that the significance of the variables can be different in micro level econometric models. Additionally, macro-economic variables are more likely correlated with each other because of their concurrences in an economic progress.

Table 3.1 Commonly Used Variable in Literature

Variable's Name*	Repeat	Underlying Paper	Paper Sample
Population	11	1-2-3-5-7-9-10-12-13-14-15	1. Dail Umamil Asri and Yoriyasu Sugie (2003) 2. Seraj Y. Abed, Abdullah O. Ba-Fail and Sajjad M. Jasimuddin (2001) 3. Md. Jobair Bin Alem and Dewan Masud Karim (1998) 4. Wenbin Wei and Mark Hansen (2006) 5. S.C. Wirasinghe and A.S. Kumarage (1998) 6. Abdullah Omer Ba-Fail (2004) 7. Dipasis Bhadra (2003) 8. Richard T. Carson, Tolga Cenesizoğlu and Roger Parker (2011) 9. Abdullah O. Ba-Fail and Seraj Y. Abed (2000) 10. Tobias Grosche, Franz Rothlauf and Armin Heinzl (2007) 11. Joyce Dargay and Mark Hanly (2001) 12. Kyung Whan Kim et al (2003) 13. Chuntao Wu and Ji Han (2011) 14. Richard A. Ippolito (1981) 15. Kopsch, F. (2012).
GDP	9	1-2-3-4-6-9-10-14-15	
Distance	5	4-7-10-13-14	
Travel Time	5	1-5-10-12-13	
GDP per Capita	4	2-6-9-11	
Price	5	4-11-13-14-15	
Service Frequency	4	1-3-4-14	
CPI	3	2-6-9	
Import Volume	3	2-6-9	
Employment	2	3-8	
Exchange Rate	2	6-9	
Cost	2	1-5	
Expenditures	2	6-9	
Fuel Price	1	8	

* The percentages are calculated out of a sample of 15 different relevant articles

As it is shown in Table 3.2, geo-economics factors are mostly related to each other. Where, values of variables are calculated by taking the product of the corresponding values for origin and destination cities. GDP values of cities are not available for the period of sample data because Bureau of State Statistics of Turkey has stopped to record city level GDP since 2001. Therefore, GDP values are estimated by assuming that GDP contribution of the cities should remain the same since 2001. In order to avoid multicollinearity problem, Bedding Capacity and Population were selected as the best representing and reliable variables amongst the other geo-economics variables since Export Volume, GDP and Intercity Migrants are all highly correlated with Population.

Table 3.2 Correlation Matrix for Geo-economics Factors

	GDP	**Bedding Capacity**	**Export Volume**	**Population**	**Intercity Migrants**
GDP	1.0000	0.1881	0.8679	0.9726	0.8246
Bedding Capacity	0.1881	1.0000	0.0773	0.1467	0.1633
Export Volume	0.8679	0.0773	1.0000	0.8680	0.6661
Population	0.9726	0.1467	0.8680	1.0000	0.8004
Intercity Migrants	0.8246	0.1633	0.6661	0.8004	1.0000

3.1 The Data

Data availability is the main issue when data coverage is to be decided. In our study, experimental model is based on the secondary source of data from 2011 for which data for all explanatory variables are available. There were 42 on-line cities in domestic network in Turkey in 2011. With this data set, we can theoretically generate 1722 different combinations of origin and destinations (O&D's). The actual number of O&D's was reduced to 861 cities by rearranging O&D's regardless of their direction. Eventually, the experimental data do not cover such number of city pairs. For some city pairs, there is no air service or the connecting flights are not meaningful due to longer travel times. In the experimental data, there are a total of 239 city pairs of which 69 city pairs are served by direct flights whereas the remaining 170 city pairs are found to be flown by connecting flights via an appropriate domestic hub. Actual data were refined by excluding the O&D's having less than 365 yearly passengers flow. Logic of this filtering was to choose meaningful connections out of the all itineraries.

Route statistics for all scheduled carriers which consist of 5 major scheduled carriers operating on domestic routes, namely Turkish Airlines, Pegasus, Atlas Jet, Onur Air and Sun Express, were used in the experimental model as a source of the dependent variable. Transfer traffic is removed from the statistics for each city pair, since the model is to estimate pure O&D passenger demand by referring to the independent variables specific to the corresponding city pairs. Airline specific data were provided by Director General Civil Aviation of Turkey (DGCA). Road distances between the cities were taken from the web site of the General Directorate Highways of Turkey. Population and bedding capacity of the cities were retrieved from the web site of State Statistics Bureau of Turkey.

Independent variables are listed in Table 3.3 where i and j stand for the cities i and j. Population, Bedding Capacity and Distance are classified as geo-economics factors having positive impact on air passenger demand. The reason for the positive impact of distance on airline passenger demand is due to the fact that the gap between relative travel times is getting higher for longer distance in favor of air transportation mode. All other independent variables are airline specific. Airlines related factors can be considered surrogate variables for airline passenger preference. Transit (as a dummy variable "O" means direct service and "1" means connecting service for a city pair), price, travel time are expected to have negative effect, whereas airline count, travel match and schedule consistency are expected to have positive effect on air passenger demand.These air service specific variables are not only driving factors for airline market share in a competitive flight route, but also showing induced effects for shifting additional customer from other modes of transportation. Schedule consistency represents seasonality of air service. Air service, which is not available for whole year, is to generate reduced passenger traffic due to lack of consistent service and consequently restricted market penetration. Airline count indicates market structure of the underlying route. Air carriers apply different business strategy depending on market form. Monopolistic airlines focus on yield improvement, whereas oligopolists seek for a better market share. These two different policies obviously influence air travel demand in opposite manner. Travel Match is reflecting round trip effect of the flight schedule for a city pair. It is simply calculated by dividing the product of number of scheduled days for outgoing and incoming week by the maximum value of the corresponding denominator which is equal to 49. According to the best knowledge of the authors, the travel match is here introduced first time in the literature to be employed in demand forecasting model for air transportation. It indicates the matching probability of a travel plan with available flights in days of week for both onward and backward direction. Travel match increases by square of the number of weekly scheduled flights, so weak frequency of service would be reasonably less attractive especially for air traveler who flies for business purposes

Table 3.3 Independent Variables

Notation	Functional Specification	Factor
	x x	Population
	x x	Bedding Capacity
	Distance by road	Distance
	Dummy variable (0 or 1)	Transit
	Average airline ticket price	Price
	Number of airlines on the route	Airline Count
	Percentage of flight choices matching travel plan	Travel Match
	Number of months that flight service is available	Schedule Consistency
	Time elapsed between start and end time of travel	Travel Time

The following independent variables are found significant to be employed in the estimation models;

3.1.1 Transit

Transit is dummy variable. It takes 0 if there is a direct flight between origin and destination city, 1 if only connecting flights are available on the itinerary. There are 4 airports which can be considered as connecting points which are İstanbul Atatürk Airport, İstanbul Sabiha Gökçen Airport, Ankara Esenboğa and İzmir Adnan Menderes. İstanbul is a multi-hub city with two airports. Each airline uses one or more hub station to base its partial or entire aircraft fleet in order to facilitate its scheduled services. Turkish Airlines uses both airports of İstanbul and Ankara Esenboğa; Pegasus uses İstanbul Sabiha Gökçen Airport; and Sun Expres uses İstanbul Sabiha Gökçen Airport and İzmir Adnan Menderes Airport; Atlas Jet and Onur Air use İstanbul Atatürk Airport as hub station of their flight operations. Since the fleet capacity is based in hub station, the majority of flights have to be performed via hub station. Network carriers such as Turkish Airlines and Pegasus are using their hubs as connecting gateway while other airlines such as Sun Express, Onur Air and Atlas Jet carry point to point traffic between their hub and spokes. For either case, cities which include hub stations are distinguished with large number of flights and high level of competition. Moreover, hub stations are usually located in a central point of wider catchment area. These hub cities such as İstanbul, Ankara and İzmir represent the top three big cities of Turkey in means of population and economic size as well.

Most of the itineraries between city pairs are not directly connected. It means that some air passengers travel with connecting flights via one or more transfer points. If there is no direct service on a city pair, the dummy variable transit becomes 1 or 0 otherwise. Obviously, passengers would less likely prefer to fly with connecting flights against direct flights or even the alternative modes of transportation. So the dummy variable transit is expected to be negatively related with air travel demand.

3.1.2 Price

Average price is available for each city pair yearly basis. Average price, which is in Turkish Lira, is adjusted by inflation rate through the years in order to make it aligned with consumer purchasing power. Airline pricing system is usually so different from alternative modes of transportation. For example, ticket prices for surface modes usually depend on travel distance, whereas for airline, ticket prices depend on market dynamics so airline ticket prices are not considered to be cost oriented.

Brons et al. (2002) pointed out that ticket price is an elastic driver for airline demand generation. So, average ticket price is employed in the model since it is considered as a stimulating factor of air travel demand.

Average ticket prices for each route and for each year are obtained from DGCA of Turkey.

3.1.3 Distance

The utility of air services would be higher when the travel distance is longer. This is due to the fact that air transportation is relatively more advantageous to road transportation for longer distance since relative travel time is significantly distinguishing by distance. Nevertheless, this proportional relation may not show up for longer range routes since long haul international routes require higher budget allocation for air traveler.

Values of city pair distances are obtained from the website of Director General Road Transportation.

3.1.4 Airline Count

Market structure is one of the key elements to describe the market demand. Turkish domestic air transport market has been deregulated in 2003; therefore, there are no barriers for any carrier to operate scheduled services on domestic routes. If there is a monopoly carrier on a route, we can interpret that the others do not find this route as attractive as the one in their current flight network.

In a monopoly market, carrier used to maximize its passenger yield and focus on profitability margin rather than seat capacity increase. Whereas, in a competitive market, carriers are focusing on market share and trying to recapture more passenger traffic by either stealing from competitor's customers or stimulating additional demand with attractive prices. So, the number of carriers on a route is supposed to be a good indicator to represent the market structure. Bigger market attracts more carriers as well as more carriers deepen the market size. Before the deregulation, there was only Turkish Airlines serving on domestic routes, now there are 5 major carriers performing scheduled services on different domestic routes.

The number of airlines operating on a route can be a good representation for the market size. The number of carriers shows the degree of competition on the route. Since the competition has direct impact on market growth, the number of carriers is also embedded in the model expecting a positive relation with air travel demand.

3.1.5 Travel Match

Travel match shows how likely a passenger finds available return flight while he is making his travel plan. If there are daily flights on a particular route then it is 100 % sure that return day would match a flight. For example, if there are 3 weekly flights, a passenger with minimum 1 week stay would find matching flight days by a probability of 9/49. If flight choices do not fit with required travel plan then passenger would likely to seek for alternative way of transportation by driving his own car or taking a bus trip.

3.1.6 Population

It is obvious that more populated cities are able to generate more air traffic. People have different reasons to travel between cities such as business, leisure or VFR (Visiting Friends and Relatives) etc. Highly populated cities are more likely to get interaction each other. There are significant numbers of intercity immigrants living in big cities. These immigrants are expected to visit their hometown in a certain period. By considering the fact that people living in urban area are more likely to travel due to more business activities and higher relative earning , urban population of the cities, which were retrieved from the website of Bureau of Statistics of Turkey, were used in the model.

Since population is related to a particular city, it is converted to city pair level by taking product of the corresponding values.

Where represents for population factor between the cities i and j,

represents for population for the city i and

represents for population for the city j.

3.1.7 Bedding Capacity

The bedding capacity of the cities is assumed to be a good indicator to represent the interactivity between the cities. Especially, leisure traffic for the purposes of VFR (Visiting Friends and Relatives) or touristic activity is considered to be a significant part of the air travel demand and is to be obviously highly related with the number of beds in the concerning cities.

There are famous touristic destinations in different regions of Turkey. People usually spend their holiday in these destinations. Bedding capacity of the cities with touristic resorts is significantly higher. In other words, the bedding capacity gives an idea about how many visitors are expected to be there during the season.

Since bedding capacity is related to a particular city, they are converted to city pair level by taking product of the corresponding values.

Where stands for bedding capacity factor between the cities i and j,

stands for bedding capacity for the city i and

stands for bedding capacity for the city j.

The bedding capacity of each city is obtained from the Ministry of Tourism.

3.1.8 Travel Time

Travel time is the duration between the start and end point of the journey. The longest domestic city pair can take up to 2 hours by air plane or 1 day by bus trip. However, total elapsed time through the connecting flights can vary depending on waiting time at transfer point. Longer connecting time adversely affect air travel demand since the gap between air and surface transportation would be indifferent while connection time is getting longer.

3.1.9 Consistency

Consistency refers to continuity of flight schedule. Airlines can change the number of weekly flight frequency or period of the flight schedule depending on size or seasonality of air travel demand on the route. Airport can be sometimes closed due to maintenance and repair facility. During this closure interval or period of unavailable service, there will be no air traffic; therefore, air travel statistics would be ceased. Connection availability can vary through the schedule terms of the year which are called as summer and winter schedule. Airline usually performs higher frequency in summer schedule; therefore, there are more connection flights in summer schedule. In case of misconnection, there will be certainly no connection traffic on underlying city pair; hence, consistency stands for an important determining factor representing air travel statistics.

3.2 Model Estimation

This research utilizes artificial neural network (ANN) in comparison with multivariate semi-logarithmic regression (MLR) model to accurately forecast air travel demand between domestic cities in Turkey. The factors which influence air traffic markets are identified, evaluated and analyzed in detail by applying the concerning models alternatively.

3.2.1 Multivariate Regression

The statistical tool SAS Jmp 10 was utilized in this research. Both multivariate semi-logarithmic regression and artificial neural network were analyzed in order to compare prediction performance.

Table 3.4. Simple Statistics

Column	N	DF	Mean	Std Dev	Sum	Minimum	Maximum
Travel Match	239	238.00	0.9150	0.2291	218.684	0.0408	1.0000
TravelTime	239	238.00	4.2922	2.7450	1025.84	0.7500	11.3000
Transit	239	238.00	0.7364	0.4415	176.000	0.0000	1.0000
AirlineCount	239	238.00	1.3640	0.8079	326.000	1.0000	5.0000
LogPrice	239	238.00	4.5162	0.1696	1079.38	4.1253	4.9867
Consistency	239	238.00	11.5241	1.5060	2754.25	2.0000	12.0000

LogPop	239	238.00	13.6928	1.4600	3272.57	10.5918	17.9776
LogBed	239	238.00	10.3268	2.1940	2468.10	3.0415	17.2165
Distance	239	238.00	859.218	379.558	205353	80.0000	1874.00

The service specific factors travel match, travel time, transit, price, airline count and consistency are embedded in the model in addition to the macro economic factors population, distance and the bedding capacity of the cities. In Table 3.4, simple statistics of the independent variables out of 239 observations are depicted where ranges of the values can be seen as well. For example, travel match varies between 4 % and 100 %. It is interesting to note that the longest distance traveled is 1.874 km and shortest period of available air service is 2 months while most of them are quite close to 12 months.

Table 3.5 shows the matrix of correlation between the independent variables. The results show that some of the variables are interrelated. For example, LogPop has a correlation coefficient of -0.7135 and -0.6591 with travel time and transit respectively. However, this relatively high correlation coefficient does not mean a certain relation of causality; on the contrary, they may concur in the specified period simultaneously without a cause and effect interaction. Furthermore, the listed Variance Inflation Factors (VIF) for coefficents of independent variables in Table 3.10 support abovementioned interpretation since all independent variables have smaller VIF value than 5, which implies that multicollinearity problem is not expected.

Table 3.5. Correlation Matrix

	Trm	**Ttm**	**Trt**	**Cnt**	**Log Prc**	**Scd**	**Log Pop**	**Log Bed**	**Dst**
Trm	1.0000	-0.3860	-0.1856	0.1663	0.1994	-0.0004	0.2999	0.2346	-0.1630
Ttm	-0.3860	1.0000	0.6528	-0.4673	-0.3203	0.0858	-0.7135	-0.5655	-0.1074
Trt	-0.1856	0.6528	1.0000	-0.2953	-0.2061	0.2118	-0.6591	-0.2795	-0.0139
Cnt	0.1663	-0.4673	-0.2953	1.0000	0.3020	0.0981	0.5363	0.4170	0.1820
LogPrc	0.1994	-0.3203	-0.2061	0.3020	1.0000	0.1329	0.3576	0.2544	0.0598
Scd	-0.0004	0.0858	0.2118	0.0981	0.1329	1.0000	-0.0691	-0.0199	0.0106

LogPop	0.2999	-0.7135	-0.6591	0.5363	0.3576	-0.0691	1.0000	0.4653	-0.1288
LogBed	0.2346	-0.5655	-0.2795	0.4170	0.2544	-0.0199	0.4653	1.0000	0.1112
Dst	-0.1630	-0.1074	-0.0139	0.1820	0.0598	0.0106	-0.1288	0.1112	1.0000

The selected predictor variables were verified via a stepwise regression analysis procedure. Table 3.6 shows that minimum corrected Akaike Information Criterion (AICc) has been reached by stepwise regression model with forward direction when all independent variables are entered into the model.

Table 3.6 Stepwise Regression

Stepwise Fit

Stepwise Regression Control

Direction : Forward

Stopping Rule : Minimum AICc

SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC
168.48174	229	0.8577459	0.8688	0.8636	10	10	617.8526	654.9307

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC
1	TravelTime	Entered	0.0000	759.1984	0.5912	478.43	2	872.383	882.71
2	AirlineCount	Entered	0.0000	192.9734	0.7415	218.14	3	764.917	778.652
3	Transit	Entered	0.0000	94.76763	0.8153	91.334	4	686.653	703.778
4	Distance	Entered	0.0000	22.95977	0.8332	62.127	5	664.421	684.918
5	LogPop	Entered	0.0000	16.75692	0.8462	41.351	6	647.074	670.924
6	Consistency	Entered	0.0000	14.28233	0.8574	23.939	7	631.269	658.454
7	LogBed	Entered	0.0060	5.900933	0.8620	17.918	8	625.601	656.104
8	LogPrice	Entered	0.0078	5.385216	0.8662	12.599	9	620.406	654.206
9	Travel Match	Entered	0.0330	3.383414	0.8688	10	10	617.853	654.931

The value of R^2 as shown in Table 3.7 is 0.86 which means approximately 86 % of the variation of air passenger demand is explained by the independent variables. In Table 3.8, adjusted R^2's are benchmarked in comparison with the relevant articles in the reference list. This comparison table shows that the studied model efficiency is relatively successful. The F test also shows that the regression model is globally significant at a level less than 0.0001.

Table 3.7 Summary of Fit

Summary of Fit		Analysis of Variance				
		Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
RSquare	0.868793					
RSquare Adj	0.863636	Model	9	1115.6081	123.956	168.4813
Root Mean Square Error	0.857746	Error	229	168.4817	0.736	**Prob > F**
Mean of Response	8.817485	C. Total	238	1284.0898		<.0001
Observations (or Sum Wgts)	239					

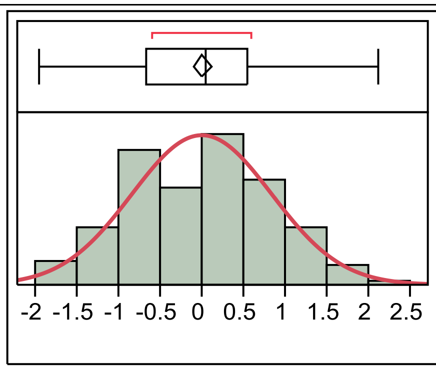
As it is shown in Table 3.8, micro level forecast (O&D) relatively underperforms macro level forecast (Aggregate) since the forecasting power lost by exploiting heterogeneous information across city pairs is dominated by the forecasting power gained due to the fact that aggregation cancels out individual forecasting errors.

Table 3.8 Model Efficiency Benchmark

Research Name	Level of Forecast	Author	Year	Independent Variables	Observation	Adjusted R Square
Demand For Air Transport In Nigeria	Aggregate	Adekunle J. Aderamo	2010	Index of Agriculture Index of Electricity GDP	26	0.923
Air Travel Domestic Demand Model in Bangladesh	Aggregate	Md.Jobair Bin Alam Dewan Masud KArim	1998	Population GDP Distance	31	0.88
An Econometric Analysis of Air Travel Demand in Saudi Arabia	Aggregate	Seraj Y.Abed Abdullah O.Ba-Fail Sajjad M.Jasimuddin	2001	Population Total Expenditures	25	0.959
Demand for Air Travel In USA	O&D	Dipasis Bhadra	2003	Density, Interaction, Distance, Marketshare, Fare	2424	0.57
An Aggregate Demand Model in Hub-and-Spoke	Aggregate	Wenbin Wei Mark Hansen	2006	Frequency, Number of Spokes, Fare, Distance, Capacity, Traffic Type	897	0.792
Gravity Model for Airline Passenger Volume Estimation	City-pairs	Tobias Grosche Franz Rothlauf Armin Heinzl	2007	Distance Population Catchment Area	956	0.761

The built MLR (Multivariate Linear Regression) model was tested for its expected normality and homoscedasticity aspects: Shapiro-Wilk W Test as shown in Figure 3.1 for normality of the residuals has shown that the null hypothesis that the data comes from a normally distributed population cannot be rejected with the p-value of 0.2256.

Figure 3.1 Normality Test



Summary Statistics

Mean	-6.32e-17
Std Dev	0.8413718
Std Err Mean	0.0544238
Upper 95% Mean	0.1072139
Lower 95% Mean	-0.107214
N	239

Goodness-of-Fit Test

Shapiro-Wilk W Test

W	Prob<W
0.992065	0.2256

The following table shows the Breusch-Pagan Test for heteroskedasticity. The p-value of the F test is 1.0 which means that we strongly fail to reject the null hypothesis for homoscedasticity in the model with the level of significance $\alpha = 100$ %.

Table 3.9 Breush Pagan Test

Summary of Fit

RSquare	0
RSquare Adj	-0.0393
Root Mean Square Error	0.857746
Mean of Response	-1.2e-16
Observations (or Sum Wgts)	239

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	9	0.00000	0.000000	0.0000
Error	229	168.48174	0.735728	**Prob >F**
C. Total	238	168.48174		1.0000

In Table 3.10, parameter estimates are depicted. As seen in the table, all independent variables are significant at 0.05 level. Since the coefficients of the logarithmic regression model represent elasticity of the corresponding independent variables, the effect of a change in any variable on the demand can be determined. For example, the price elasticity of air passenger demand as shown in Table 3.11 is approximately -1.05 which implies that if price decreases by 1%, air passenger demand would increase by 1.05%. This means that domestic air travel demand shows a slightly elastic behavior in Turkey.

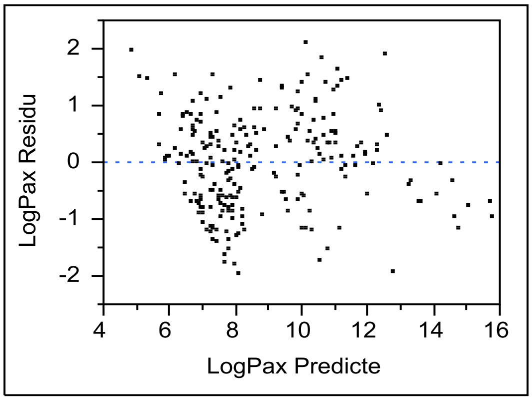
Table 3.10 Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	5.0953283	1.808305	2.82	0.0053	.
Travel Match	0.5854434	0.273002	2.14	0.0330	1.2649877
TravelTime	-0.146405	0.0363	-4.03	<.0001	3.2119283
Transit	-1.893718	0.188805	-10.03	<.0001	2.2478437
AirlineCount	0.9592384	0.088819	10.80	<.0001	1.6655125
LogPrice	-1.050123	0.363063	-2.89	0.0042	1.226691
Consistency	0.1878425	0.038815	4.84	<.0001	1.1054137
LogPop*	0.3270246	0.068732	4.76	<.0001	3.2572982
LogBed*	0.0934883	0.032053	2.92	0.0039	1.5998215
Distance	0.0012053	0.000166	7.28	<.0001	1.2778102

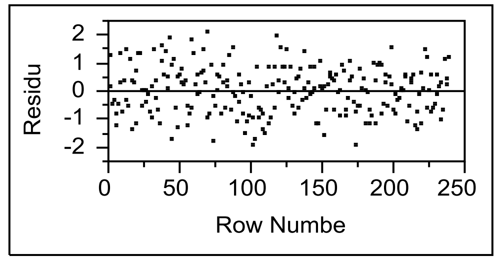
In Figure 3.2, model fit of the experimental data is shown as a scatter diagram. The scatterplot and fitted line in the figure suggest a linear pattern. The fitted values of Log (Pax) are on the horizontal axis and the residuals are on the vertical axis. Points are scattered evenly above and below the horizontal line at 0, without any sign of curvature. The residual plot does not show that the variance of the error terms change in systematic ways with the fitted values.

Figure 3.2 Residual Plots

By Predicted Value



By Row Number



In Table 3.11, the prediction expression is given. Price elasticity of demand implies that domestic air passenger traffic grows higher than the decreasing rate of ticket price. Hence, low fare policy is generally recommended for those air carriers which seek for higher load factor on a flight route since total revenue is expected to increase inspite of lower price as a result of elastic demand.

Table 3.11 Prediction Expression

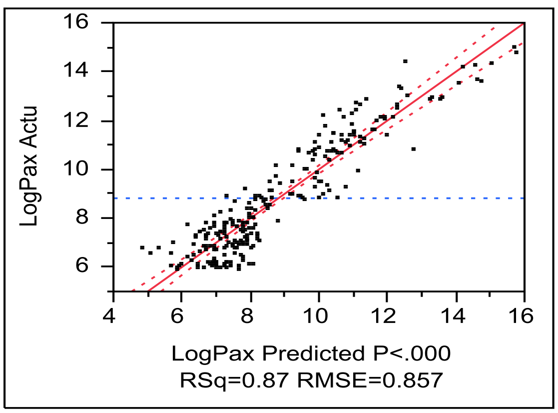
Log( ) = 5.095 + 0.585x - 0.146x - 1.893x + 0.959x -
1.050xLog) + 0.188x + 0.327x) + 0.093xLog)* + 0.001x

It can be observed that all independent variables have positive impact on air passenger demand estimation, except price, travel time and transit. The coefficient of the dummy variable transit implies that if any city pair does not include a direct service, air passenger demand would be 189 percentage point less likely than the city pair with direct service. This indicates that availability of direct air services is noticeably stimulating passenger demand on city pairs.

Another result is that airlines competition in each O&D market is significantly contributing to the air transportation demand. Competition leads to variation in prices, and consequently, consumer surplus reaches higher level once airlines try to attract more passengers by lowering prices. Price war and more frequent flights are the key factors to enlarge the market size because low fare policy enables airlines to stimulate air passenger demand in addition to shifting more passengers from other modes of transportation while high frequency of flight service attracts more air traveler.

In the figure 3.3, model fit of the experimental data is shown in scatter diagram.

Figure 3.3 Actual by Predicted Plot



3.2.2 Artificial Neural Network

This study focuses on the parallel assessment of neural network and linear regression model's forecasting performance, so the same inputs and the same data were used for the development and comparison of the two approaches. Nodes in the input layer represent independent variables of the problem, including the geo-economics and airline specific factors used in semi-logarithmic model.

The prediction performance of a neural network makes it a good alternative classification and forecasting tool in business application. In addition, a neural network is expected to be superior to traditional statistical methods in forecasting since a neural network is better able to recognize the high-level features, such as the serial correlation, if any, of a training set. Another advantage of applying a neural network to forecasting is that a neural network can capture the non-linearity of samples in the training set.

Demand forecasting for domestic air transportation in Turkey, measured by the number of passengers traveled on a domestic flight route in both direction, can be stated as:

Log ( ) = f (, , , ,Log), ,), Log),),

Where

	=	The number of passengers for the city pair i and j,
		Travel match for the city pair i and j,
		Travel time for the city pair i and j,
	=	Transit for the city pair i and j,
	=	The number of airline for the city pair i and j,
Log	=	Logarithm of price for the city pair i and j,
	=	Schedule consistency for the city pair i and j,
	) =	Logarithm of population factor for the city pair i
Log	) =	Logarithm of bedding capacity factor for the city
	=	Distance for the city pair i and j.

Air travel demand for the city pair was calculated in the following way: Neural Network model adopts the data for network training and validation. Available data set was divided into two subsets. The first was included data from the year 2011 and used for ANN training, and the second subset was the year 2010, which was absolutely unknown to the model and used for the evaluation of the forecasting ability of the developed ANN. The same procedure was previously followed in the case of MLR model.

The nine nodes in the input layer contained the following variables:

Travel match, travel time, distance, airline count, LogPrice, LogPop, LogBed, transit, and consistency.

The output variable used in the model LogPax represents the number of domestic passengers for a city pair in Turkey. Fig 3.4 shows the neural network, with two hidden layers, which was used in this research to determine the number of domestic passengers for a city pair in Turkey.

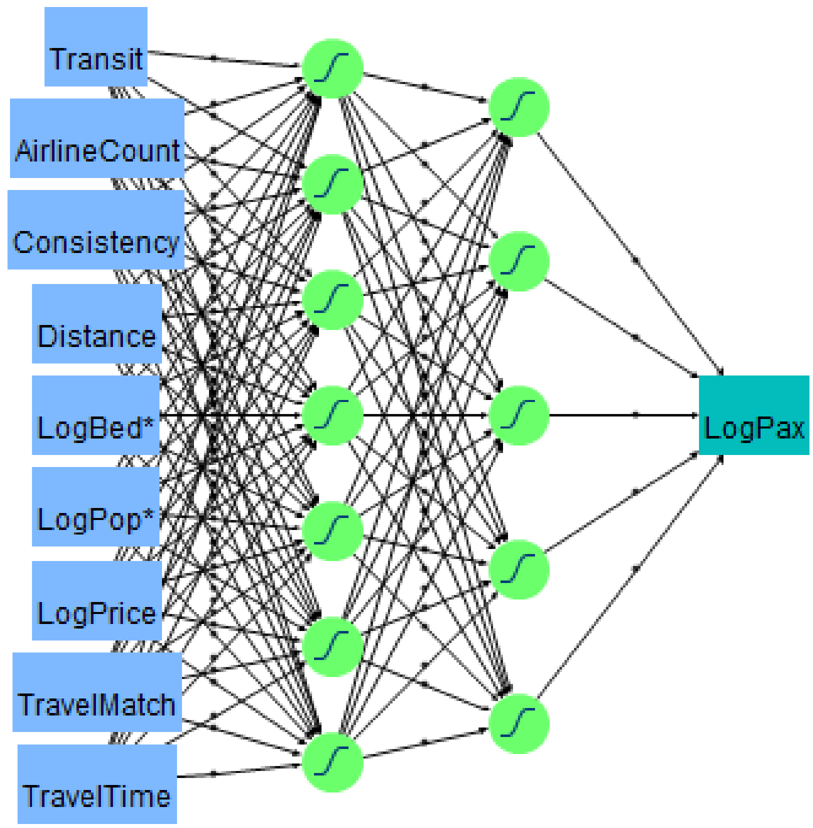
A computer program is developed using SAS Jmp version 10 for the neural network model exhibited in Fig 3.4. There are nine input nodes (independent variables) and one output node (dependent variable). Zhang (1994) finds that networks with two hidden-layers can model the underlying data structure and make predictions more accurately than one hidden layer networks. There is no standard formula to calculate the number of nodes needed in the hidden layer (Wang & Sun, 1996). The most common way in determining the number of hidden nodes and the number of hidden layers is via experiments or by trial-and-error. In this research, two hidden layers were found to generate better prediction performance where 5 hidden nodes exist in the first; and, 7 hidden nodes exist in the second hidden layer.

Back-propagation neural network (BPN) technique was used to formulate demand forecasting model for domestic air transportation in Turkey. BPN is a layered feed-forward supervised network which is suitable to establish air passenger forecasting models. The most popularly used training method is the back propagation algorithm which is essentially a gradient steepest descent method (Dougherty 1995). For the gradient descent algorithm, a step size, which is called the learning rate in ANNs literature, must be specified. The learning rate is essential for back propagation learning algorithm since it determines the magnitude of weight changes. It is well known that the steepest descent suffers the problems of slow convergence, inefficiency, and lack of robustness. Furthermore, it can be very sensitive to the choice of the learning rate. Smaller learning rates may cause network oscillation in the weight space.

The network architecture is also characterized by the interconnections of nodes in layers. The connections between nodes in a network fundamentally determine the behavior of the network. Similar to the most forecasting models as well as other applications, the networks, as shown in Table 3.12 and Table 3.13, are fully connected in that all nodes in one layer are only fully connected to all nodes in the next higher layer except for the output layer.

Each link between nodes has an associated weight, which represents the strength of the connection. Three steps identify the operation of a node, including receiving input signals and connection weights, summing up the information, and transforming it by the activation function to produce an output signal. The activation function used in the study is the widely used hyperbolic tangent function.

+Figure 3.4 ANN Model Diagram



In the training process, a large number of examples (239 observations) belonging to 2011 were employed. Using the current weights, BPN computes a set of outputs with the example input data. These network outputs are then compared against their corresponding values in the example set by computing the sum of square error. After that, the weights are updated by a partial derivative function, which propagates the errors back to the input layer.

Table 3.12ANN Model Estimates (First Hidden Layer)

Parameter	Estimate	Parameter	Estimate
H1_1:H2_1	1.556136	H1_3:Intercept	0.381731
H1_1:H2_2	0.591699	H1_4:H2_1	0.689904
H1_1:H2_3	-0.2735	H1_4:H2_2	-0.56792
H1_1:H2_4	-1.76949	H1_4:H2_3	1.253734
H1_1:H2_5	0.990594	H1_4:H2_4	-0.88014
H1_1:H2_6	-0.19781	H1_4:H2_5	0.358388
H1_1:H2_7	0.836439	H1_4:H2_6	-0.49074
H1_1:Intercept	0.242277	H1_4:H2_7	-0.90841
H1_2:H2_1	1.034874	H1_4:Intercept	0.876904
H1_2:H2_2	-0.4365	H1_5:H2_1	0.324272
H1_2:H2_3	0.143751	H1_5:H2_2	-0.00211
H1_2:H2_4	0.55859	H1_5:H2_3	0.295065
H1_2:H2_5	-0.31335	H1_5:H2_4	0.244044
H1_2:H2_6	0.195736	H1_5:H2_5	0.250195
H1_2:H2_7	1.167627	H1_5:H2_6	-0.19883
H1_2:Intercept	0.959745	H1_5:H2_7	-0.38546
H1_3:H2_1	0.139644	H1_5:Intercept	-0.14124
H1_3:H2_2	-0.1954	LogPax_1:H1_1	4.235959
H1_3:H2_3	-0.12813	LogPax_2:H1_2	-3.23628
H1_3:H2_4	-0.06562	LogPax_3:H1_3	-0.8823
H1_3:H2_5	0.006251	LogPax_4:H1_4	-3.5458
H1_3:H2_6	-0.0647	LogPax_5:H1_5	1.070927
H1_3:H2_7	0.283196	LogPax_6:Intercept	11.01466

If the operation in the learning process is successful, the global error reduces gradually, and the process eventually reaches a convergent result. The parameter estimates for weights for the first layer nodes are shown in Table 3.12; the parameter estimates of weights for the second layer nodes are shown in Table 3.13.

Table 3.13 ANN Model Estimates (Second Hidden Layer)

Parameter	Estimate	Parameter	Estimate
H2_1:Transit	-0.55373	H2_4:LogPop	-0.39156
H2_1:AirlineCount	-0.37536	H2_4:LogPrice	1.617446
H2_1:Consistency	0.572464	H2_4:RT_Choice	-1.59762
H2_1:Distance	-0.0023	H2_4:TravelTime	0.005914
H2_1:LogBed	-0.23659	H2_4:Intercept	8.462845
H2_1:LogPop	-1.02748	H2_5:Transit	-0.84658
H2_1:LogPrice	-5.24669	H2_5:AirlineCount	-0.10006
H2_1:RT_Choice	-3.12028	H2_5:Consistency	0.046918
H2_1:TravelTime	-0.30815	H2_5:Distance	-0.00178
H2_1:Intercept	41.26767	H2_5:LogBed	0.064442
H2_2:Transit	-0.21902	H2_5:LogPop	0.3272
H2_2:AirlineCount	0.084848	H2_5:LogPrice	-6.13625

H2_2:Consistency	-0.52869	H2_5:RT_Choice	-2.30182
H2_2:Distance	-0.00046	H2_5:TravelTime	0.438036
H2_2:LogBed	0.019276	H2_5:Intercept	24.86766
H2_2:LogPop	0.014382	H2_6:Transit	0.268117
H2_2:LogPrice	3.201557	H2_6:AirlineCount	-0.05302
H2_2:RT_Choice	-0.08158	H2_6:Consistency	-0.09729
H2_2:TravelTime	-0.1021	H2_6:Distance	-0.00115
H2_2:Intercept	-7.82197	H2_6:LogBed	-0.09377
H2_3:Transit	1.606374	H2_6:LogPop	0.081717
H2_3:AirlineCount	-1.27938	H2_6:LogPrice	1.900769
H2_3:Consistency	-0.09893	H2_6:RT_Choice	-0.01247
H2_3:Distance	-0.00346	H2_6:TravelTime	-0.11192
H2_3:LogBed	-0.24413	H2_6:Intercept	-6.43585
H2_3:LogPop	0.389642	H2_7:Transit	-0.33976
H2_3:LogPrice	-2.34216	H2_7:AirlineCount	0.17369
H2_3:RT_Choice	-2.57452	H2_7:Consistency	0.297075
H2_3:TravelTime	0.441362	H2_7:Distance	-0.0005
H2_3:Intercept	13.18874	H2_7:LogBed	-0.13734
H2_4:Transit	-0.05905	H2_7:LogPop	0.346563
H2_4:AirlineCount	-0.30277	H2_7:LogPrice	3.589626
H2_4:Consistency	-0.48702	H2_7:RT_Choice	-0.05449
H2_4:Distance	-0.00166	H2_7:TravelTime	-0.13956
H2_4:LogBed	-0.10312	H2_7:Intercept	-21.7768

Training set is used to adjust the weights on the neural network. Whereas, validation set is used to minimize over fitting. Although the weights of the network with validation data set were not adjusted, they are used to verify that any increase in accuracy over the training data set actually yields an increase in accuracy over a data set that has not been shown to the network before or at least the network has not trained on it. If the accuracy over training data set increases while the accuracy over validation data set stays the same or decreases, the neural network is over fitting so training should be stopped.

Training set composes of 2011 data which includes 239 observations. Validation set is randomly (with the probability ≤ 0.5) generated from 2010 data of which the remaining sample is selected for test data. Table 3.14 shows that ANN model with R^2 of 0.955 fits training set quite well, significantly better than semi-logarithmic regression model. Nevertheless, ANN model shows a good prediction performance with R^2 of 0.917 on test data as well. It is interesting to say that prediction performance of ANN model on test data is better than the prediction performance of semi-logarithmic regression model on its training set in means of R^2 statistics.

Table 3.14 ANN Summary of Fit

Training		Validation		Test	
LogPax	**Measures**	**LogPax**	**Measures**	**LogPax**	**Measures**
RSquare	0.9552903	RSquare	0.925799	RSquare	0.9171494
RMSE	0.490117	RMSE	0.6739678	RMSE	0.6752319
Mean Abs Dev	0.3861082	Mean Abs Dev	0.5326967	Mean Abs Dev	0.5561586
-LogLikelihood	168.69277	-LogLikelihood	111.65586	-LogLikelihood	106.7289
SSE	57.411317	SSE	49.51136	SSE	47.417561
Sum Freq	239	Sum Freq	109	Sum Freq	104

Table 3.15 shows prediction expression for the output node which is the solution of ANN model. In the first hidden layer, there are 5 hidden nodes. Therefore, the prediction expression composes of the estimated weights of those 5 connections from the first hidden nodes to the output node together with the biased parameter.

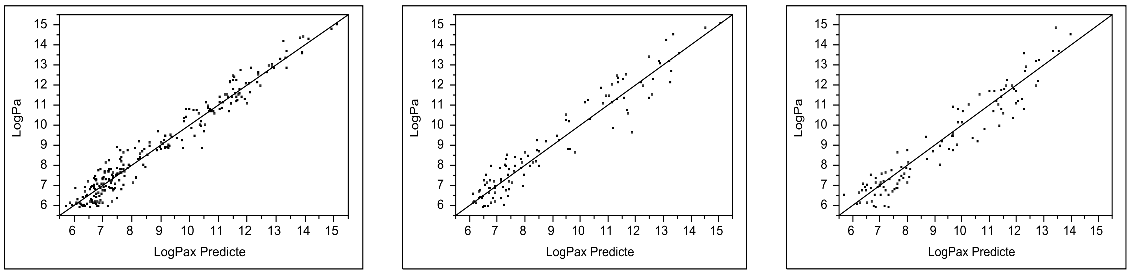
Table 3.15 ANN Prediction Expression

LogPax = 11.0146605286637 + 4.2359587194418 H1_1 + -3.23628110540887 H1_2 + -0.882302675958656 H1_3 + -3.54579520992574 H1_4 + 1.07092679366109 H1_5

Figure 3.5 shows the visual representations for the model fitness with predicted values over available sample data. The prediction values for each set of data consistently fall along a linear pattern signifying that the predicted values are similar to the actual values through the given set of data interval.

Figure 3.5 ANN Actual versus Prediction

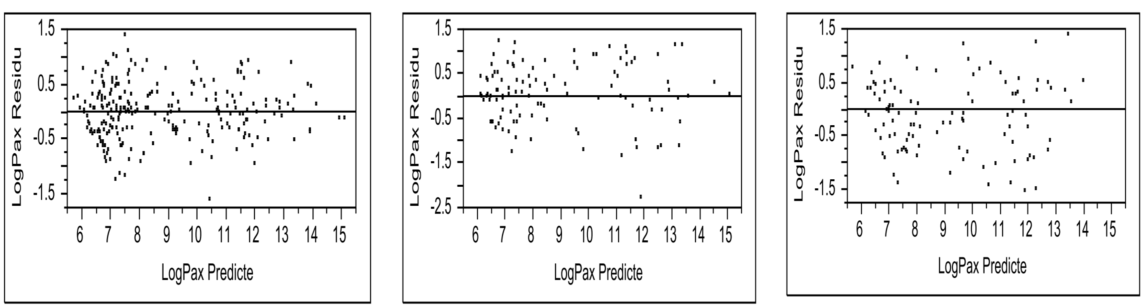
Training **Validation** **Test**



In Figure 3.6, residual plot of each set of data is shown as a scatter diagram. The fitted values of Log (Pax) are on the horizontal axis and the residuals are on the vertical axis. Points are scattered evenly above and below the horizontal line at 0, without any sign of curvature. The residual plot does not show that the variance of the error terms change in systematic ways with the fitted values.

Figure 3.6 ANN Residual Plot

Training **Validation** **Test**



3.3 Prediction Performance

A fundamental concern in forecasting is the measure of forecasting error for a given data set and a given forecasting method. Accuracy can be defined as “goodness of fit" or how well the forecasting model is able to reproduce data that is already known (Makridakis and Wheelwright, 1989).

This study used three standard error measures: R square (R^2), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

1. Mean Absolute Percentage Error (MAPE):

The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of accuracy of a method for constructing fitted values in prediction model. It usually expresses accuracy as a percentage, and is defined by the formula:

where A_i is the actual value and F_i is the forecast value.

The difference between A_i and F_i is divided by the actual value A_i again. The absolute value in this calculation is summed for every fitted or forecasted point in time and divided again by the number of fitted points n. Multiplying by 100 makes it a percentage error.

2. Root Mean Squared Error (RMSE):

RMSE is a measure of dispersion of forecast errors, statisticians have taken the root of the average of the squared individual errors. The smaller the RMSE value, the more accurate the model. The following equation describes the root mean squared error measurement:

3. R Square Statistics (R^2):

R^2 is generally a statistic that will give some information about the goodness of fit for a model. An R^2 of 1.0 indicates that the regression line perfectly fits the data. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression line approximates the real data points.

Where, =

.

A test data was randomly selected out of 2010 traffic results consists of 213 observations. In Table 3.16, it is shown that the MLR model predicts test set of 2010 figures with a MAPE (Mean Absolute Percentage Error) value 9.0 % and a RMSE (Root Mean Square Error) value of 0.8184.Whereas, ANN model predicts the same test data with a RMSE of 0.6752 and a MAPE value of 6.4 %. R^2's of ANN model are also significantly better than those of MLR Model.

Table 3.16 Model Performance Comparison

Data Set	Year	N Rows	MLR			ANN		
			R^2	MAPE	RMSE	R^2	MAPE	RMSE
Training	2011	239	0.8688	8.3 %	0.8577	0.9553	4.7 %	0.4901
Validation	2010	109	0.8722	8.1 %	0.8845	0.9258	6.1 %	0.6740
Test	2010	104	0.8184	9.0 %	0.9997	0.9171	6.4 %	0.6752

In spite of significant fluctuation by 16 % in number of passengers between the years 2010 and 2011, prediction performance of both models seem to be pretty stable in addition to that ANN model shows more reliable prediction performance in comparison with MLR model.

Air travel demand estimated by a neural network is very close to the actual values. In other words, the forecasting output from a neural network is accurate, with a relatively small amount of error. The low MAPE indicates that the deviations between the predicted values derived by the neural network and the actual values are very small.

After completion of an estimation models, the MAPE values were compared between ANN and MLR models to appraise the performance of the two alternative prediction models. These standards are defined as Table 3.17 where denotes the actual value, denotes the predicted value . The lower the MAPE values, the more accurate the prediction. Lewis (1982) presented the standard levels of MAPE (%) as shown in Table 3.17.

Table 3.17 explains the best conclusion (<10 %) for ANN model evaluation that is highly accurate forecasting in these experiments.

Table 3.17 The Standard Levels of MAPE model evaluation

< 10 %	Highly accurate forecasting
10-20 %	Good forecasting
20-50	Reasonable forecasting
>50 %	Weak and inaccurate predictability
Criteria of MAPE (%) Forecasting ability Source: Lewis	

Not only ANN with two hidden layers outperforms the multiple regression models in forecasting accuracy but also it shows highly acceptable range of prediction performance.

MAPE is a relative assessment used for comparison across the testing data because it is easy to interpret, independent of scale, reliable and valid (Rob Law et al., 1999).

In order to compare hit rate of both models, Z is defined as a relative measurement for acceptance level. As a reference rate for optimal experimental outcome, Z was set for ± 10 % in this assessment.

, for

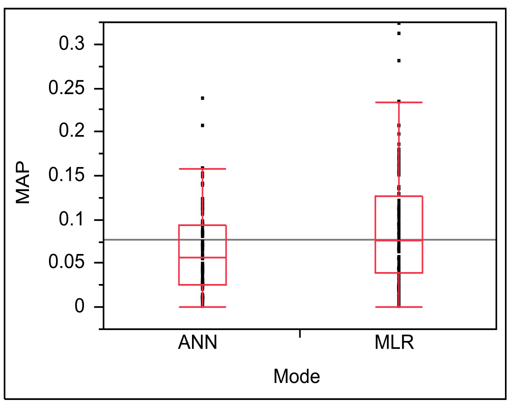
Where and represent the forecasted and actual number of passenger in logarithmic form for i=1 to i=104, respectively. In Table 3.18, it is shown that within a 10 % discrepancy range, neural network succeeds in achieving 79 % of output in the acceptable range. Hit rate significantly improves to 93 % for neural network model when city pairs are restricted by hub and spokes. In other words, both models are more successfull to accurately predict air passenger demand for the flight routes which include one of the three hub cities İstanbul, Ankara and İzmir. These hub and spoke routes are classified by high-profile geo-economics factors in addition to their competitive market structure. Although neural network model relatively outperforms a better accuracy, prediction performance of both models for spoke to spoke routes are weaker. The fact that airlines expedite potential traffic from hub and spoke routes in a cultivating way may be a good reason to explain such differentiation in prediction performance.

Table 3.18 Hit Rate Comparison

Z Values	ANN	MLR
All Sample	79 %	63 %
Hub and spokes	93 %	73 %

In order to compare two sample means, Wilcoxon test is conducted since distributions of MAPE values are not even close to looking normal. Wilcoxon test is indeed the equivalent of another well known nonparametric test as the Mann-Whitney U Test.

Figure 3.7 Oneway Analysis of MAPE by Model



Wilcoxon Tests (Rank Sums)

Level	Count	Score Sum	Expected Score	Score Mean	(Mean-Mean0)/Std0
ANN	104	9773.00	10868.0	93.971	-2.522
MLR	104	11963.0	10868.0	115.029	2.522

2-Sample Test, Normal Approximation

S	Z	Prob>|Z|
11963	2.52174	0.0117*

1-way Test, ChiSquare Approximation

ChiSquare	DF	Prob>ChiSq
6.3650	1	0.0116*

Oneway analysis of MAPE values of both models is shown in Figure 3.7. P-values, which are less than 0.05, conventionally implies that the null hypothesis, the distributions of the both groups are centered in the same place, is rejected. Therefore, based on these two MAPE samples of ANN and MLR, it is concluded that prediction performance of ANN model is significantly better than prediction performance of MLR model by MAPE assessment.

3.4 Sensitivity Analysis

The coefficients of multivariate regression model have partial effect, or ceteris paribus, interpretations. In other words, sensitivity analysis of one independent variable can be easily conducted by holding other independent variables fixed. The advantage of multiple regression analysis is that it provides this ceteris paribus interpretation even though the data have not been collected in a ceteris paribus circumstances.

The ability to estimate the effect of an independent variable independent of the other independent variables in the model is a very powerful and compelling aspect of regression.

Logarithmic regression is also called the constant elasticity model because the coefficients can be interpreted as the percentage change in dependent variable for a 1 percent change in independent variable. Moreover, in this type of model the elasticity is constant and does not change over the range of corresponding independent variable. It is constant at the level of the estimated coefficient.

Concerning the multivariate semi-logarithmic model, it is found that all nine independent variables contribute significantly. All independent variables have a positive relationship with air passenger demand except price, travel time and transit which are related to either level of service quality or purchasing power of customer. While number of connections or journey time are increasing, air passenger demand gets smaller as expected. Price variation has direct influence on purchasing power of customer. Those people with restricted budget could not afford higher price of air ticket, instead they are expected to prefer cheaper alternative modes of transportation. Therefore, high yield policy would reduce potential market size , resulting in less number of target customers who can afford to fly with even relatively expensive prices.

Since multivariate semi-logarithmic regression model is linear, price elasticity of air passenger demand remains constant regardless of the values of any of the other variables in the model. For small changes in price, price elasticity of air passenger demand can be formulated as follows;

Where = Price elasticity of air passenger demand,

D = Air passenger demand,

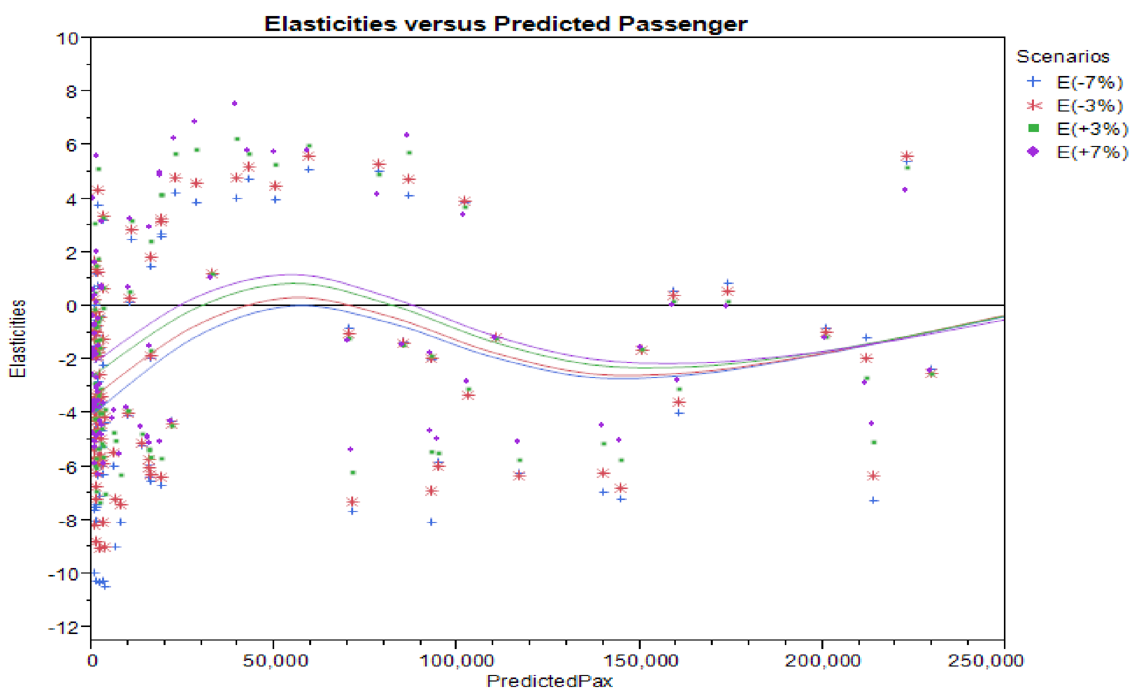
p = Air ticket price and

= Coefficient of log (price) in MLR model.

With this approximation, MLR model implies that . It means that price elasticity of air passenger demand in Turkey is slightly elastic.

On the other hand, ANNs are often considered to be “black boxes” with little capacity to provide insight on the dataset from which they have been generated. Since ANN presents non-linearities, the elasticity of any one variable against any other depend on the values which the variables have. In figure 3.8, point elasticity of air passenger demand is shown against predicted pax, using test data by ANN. Point elasticities are estimated by ANN model for 4 different scenarios of price changes with constant percentages. Point elasticity fluctuates from point to point through the range of predicted air passenger. It even oscillates up and down, resulting in having opposite sign in small interval. Therefore, ANN model seems to be unable to explain overall effect of price variation. Because of nonlinearity aspect of the neural network, the output may have many different sensitivities over the full range of a particular input. Thus, unlike classic statistical models, in a network it does not appear to be easy to find out the effect that each explicative variable has on the dependent variable (Montano, 2003). However, in general, neural networks are applied in cases where the predictive accuracy is of greater importance than understanding the learned concept. Nevertheless, ANN model is capable of better estimating point elasticity at a specific situation within an input domain since ANN is more precisely able to predict incremental effect of an input on the output.

Figure 3.8 Point Elasticity of Demand by ANN



CHAPTER 4 IMPLICATIONS AND APPLICATIONS

Demand forecasting models can be applied in various areas of aviation industry. Micro level forecasting with true O and D's brings many advantages in means of detail evaluation of each individual market. The amounts of true O and D passengers, points to point traffic, in addition to beyond traffic which can be served via network hub are taken into account by airlines manager for flight scheduling and network structuring processes. There are many decision items depending on air passenger demand forecast by O and D level such as aircraft capacity, flight frequency and connecting flights etc. It is obvious that accuracy in air passenger demand forecasting would highly support airline decision makers in order to attain efficient network management and its route economies. O and D level forecast can be aggregated to reach macro level forecast which is more meaningful for airport operators or government agencies since they focus on total traffic of an airport, a region or nationwide aviation industry.

In the following sections where air passenger demand is forecasted for potential new airports or new routes, geo-economics and airline service related factors, based on 2011 figures, are calculated by these assumptions:

1. Scheduled flights are consistently available year around.
2. Market structure is monopoly.
3. Travel match is 100% which implies that services are available daily basis.
4. Average price is 90 TL.
5. Routes are served by direct service which implies that the dummy variable transit takes 0 values.

6. Travel time is derived from aircraft block hours of underlying route.

4.1 Potential Cities for New Airport

Highest potential destination which is expected to firstly start service from a new airport would be most likely one of the hub cities such as İstanbul or Ankara. Prediction expression generates the largest number of passengers for the hub cities İstanbul and Ankara since they both have the highest profile in means of model input variables. In addition to that they are both able to connect with beyond domestics destinations because of large number of connecting flights via their airports. Therefore, predicted passenger values from each potential new airport to the hub cities are listed as given in Appendix D. The city pairs having a less amount of distance than 300 km are considered inconvenient since short distances are not suitable for air transportation choice. Moreover, total predicted passenger numbers, which are considered to be potential connecting passengers in case of a new air service between the new airport and a hub station, are listed in Appendix D.

In order to compare two alternative hubs and reflect the ranking between the spoke airports, detour factors are averaged for the itineraries in domestic flight network by weighting them with passenger number.

Let's assume O, H and D represent for Origin Airport, Hub Airport and Destination Airport respectively. Then,

Where, DF = detour factor,

|OH| = distance between O and H,

|HD| = distance between H and D,

|OD| = distance between O and D.

Detour factor shows the ratio between the connecting distance and direct distance between the city pair. So, small values of detour factor indicate the acceptable range of transfer traffic against extra connection time. There may be beyond destination city with higher predicted number of passenger. However, if a high value of detour factor belongs to underlying destination cities then it would be less likely acceptable for the passengers due to longer elapsed time. Therefore, Weighting Average Detour Factor (WADF) is calculated by passenger numbers to help making the alternative hubs to be comparable by giving higher weight to the city pair with bigger passenger flow.

Weighted Average Detour Factor (WADF) is calculated for the airport A_i as follows;

Mersin predominantly takes first place as a potential new airport. The biggest potential destination for Mersin would be İstanbul at an average 1,011 non-directional passengers for each day. This estimated daily traffic requires at least daily 3 flights on the route by a typical narrow body aircraft capacity. The highest total number of potential passengers for all other domestic cities, with an estimated total number of yearly passengers of 1,942,631 is again for Mersin. There are some potential airports of which yearly passenger traffic are more than the current statistics of some domestic airports in

Turkey as shown in Appendix E. This result raises a question about reliability of decision criteria for new airport investment.

4.2 Potential City Pair for New Routes

Potential new routes between available airports were depicted in Appendix C. Only top 79 potential routes were listed in the table. The potential route Bursa and Antalya takes first place with estimated yearly passengers of 125,168 which imply weekly 9 flights by small narrow body aircraft on the route. Most of the listed routes require less frequency than daily. However, utility of air transport mode is more effective if there is enough number of flight choices like daily or even multiple departures at prime times of the day.

4.3 Evaluation of Hub Cities for Connectivity

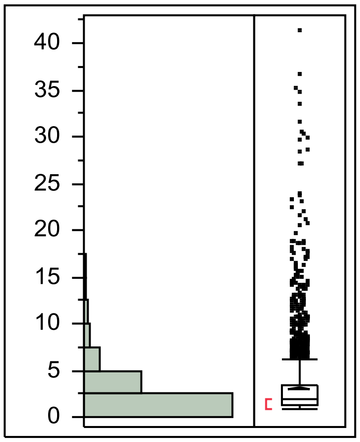
Figure 4.1 shows Ankara is better hub with the smallest WADF 1.33. 50 % of all city pair combinations via Ankara have a detour factor of less than 1.27 which is quite acceptable for connecting passengers. This result is not surprised since the capital city Ankara is geographically located in central point of Turkey.

There is a significant difference between average detour factor and WADF for İstanbul. It implies that the itineraries with higher traffic flow consist of relatively lower value of detour factor via İstanbul so that it cancels out the negative impact of its location disadvantage in domestic flight network.

Detour factor of İzmir as a domestic hub candidate is similar to the detour factor of İstanbul. High passenger traffic flow via İzmir enables it to compensate location disadvantage of İzmir.

Figure 4.1 Detour Factors of Potential Hubs

Hub = İstanbul



Quantiles

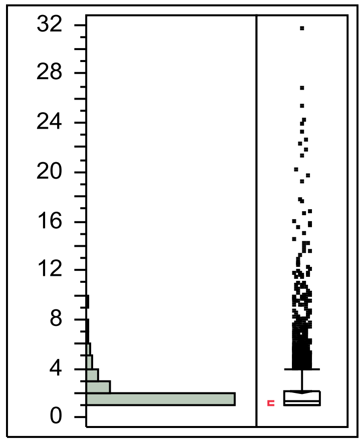
100.0%	maximum	41.2727
99.5%		23.6079
97.5%		12.5414
90.0%		6.0239
75.0%	quartile	3.36892
50.0%	median	2.00958
25.0%	quartile	1.40689
10.0%		1.0848
2.5%		1
0.5%		1
0.0%	minimum	1

Summary Statistics

Mean	3.1159424
Std Dev	3.4205858
Std Err Mean	0.0600936
Upper 95% Mean	3.2337676
Lower 95% Mean	2.9981171
N	3240

WADF = 1.59

Hub = Ankara



Quantiles

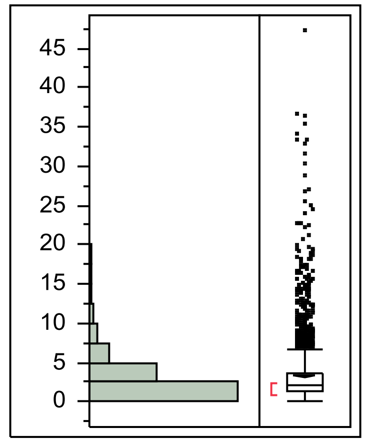
100.0%	Maximum	31.6667
99.5%		16.6943
97.5%		8.57091
90.0%		4.09177
75.0%	Quartile	2.1825
50.0%	Median	1.27559
25.0%	Quartile	1.0286
10.0%		1
2.5%		1
0.5%		1
0.0%	Minimum	1

Summary Statistics

Mean	2.1380508
Std Dev	2.3682298
Std Err Mean	0.0416056
Upper 95% Mean	2.2196266
Lower 95% Mean	2.0564749
N	3240

WADF = 1.33

Hub = İzmir



Quantiles

100.0%	maximum	47.1591
99.5%		24.3772
97.5%		13.5972
90.0%		6.38886
75.0%	quartile	3.58391
50.0%	median	2.16465
25.0%	quartile	1.44297
10.0%		1.12873
2.5%		0.87436
0.5%		0.50186
0.0%	minimum	0.18502

Summary Statistics

Mean	3.2758798
Std Dev	3.5740647
Std Err Mean	0.0627899
Upper 95% Mean	3.3989918
Lower 95% Mean	3.1527678
N	3240

WADF = 1.70

4.4 Design of a New Hub and Spoke Network

Airline hub is an airport that an airline uses as a connecting point to get passengers to their final destination. It is part of a hub and spoke network, where air travelers moving between airports not served by direct flights change planes en route to their destinations. This is as opposed to the Point to Point model.

Table 3.18 shows that ANN model is more successful in predicting air passengers for hub and spoke routes. Therefore, predicted passengers by ANN can be also useful to develop a new hub and spoke structure of a regional flight network out of spoke destinations. One of the most important selection criteria for a new hub airport is that the total point to point air passenger demand originated or destined by a hub airport should be on the top of the candidate list. Table 4.1 shows air passenger demand for all airport pairs in Mediterranean Region of Turkey. Antalya has the highest bilateral traffic within the airports in the region. Detour factor, weighted by predicted air passengers, for Antalya is 1.17 which is much better than weighted detour factor for Adana which is 4.99. These implies that Antalya is the best hub alternative in Mediterranean Region. Moreover, Adana-Isparta and Hatay-Isparta have significant air passenger demand, therefore hub and spoke model should have them connected via Antalya.

Table 4.1 Hub Selection in Mediterranean Region

	Adana	Antalya	Hatay	Isparta	Kahramanmaraş	Sum
Adana		112.335	800	32.887	891	146.913
Antalya	112.335		4.275	30.063	3.407	150.080
Hatay	800	4.275		26.344	10.763	42.182
Isparta	32.887	30.063	26.344		21.671	110.965
Kahramanmaraş	891	3.407	10.763	21.671		36.732
Sum	146.913	150.080	42.182	110.965	36.732	486.872

CHAPTER 5 CONCLUSION

5.1 Results and Discussions

The developed ANN model consists of two hidden layers with 5 and 7 artificial neurons respectively. The choice of both the type and the architecture of the structure of the developed ANN prognostic model were done by applying the trial and error method. In order to train the ANN model, the independent variables (Table 3.3) of geo-economics and service related factors were used as input data. The same independent variables were chosen as input training data in order that the results of the developed ANN model are comparable with those of the MLR developed model. The ANN output is the dependent variable which is the logarithm of the number of passenger for a city pair i and j. In other words, the input training data are the independent variables of the developed MLR forecasting model, as well as the output of the ANN model is the dependent variable of the developed MLR forecasting model.

As mentioned before, the used data set was divided into two subsets. The first was included data from the year 2011 and used for the ANN training, and the second subset was out of the year 2010, which was absolutely unknown to the model and used for the evaluation of the forecasting ability of the developed ANN. The same procedure was followed in the case of developed MLR model.

Table 3.15 shows the prediction of all the developed model is in a very satisfactory level (MAPE $\leq$ 10 %). In summary, prediction accuracy is most reliable by using the ANN approach compared to the forecasting by using MLR model.

Macroeconomic factors such as urban population, distance and the number of beds in tourism facility are found to be prominent drivers for domestic air transport demand. However, geo-economics progress is only unable to explain such significant increase in airline industry. Therefore, airline service related factors in conjunction with its business strategy and practices seem to be distinguishing drivers for fluctuation in air travel demand. However, actual air passenger traffic may not go in parallel with potential air travel demand on a flight route. In other words, product and services supplied by airlines may be insufficient to cultivate market potential in depth. For example, if flight choices are not wide, or bidirectional connections are missing, or ticket prices are high due to monopolistic market situation then it would not be quite possible to convert potential customer into actual passenger. This fact may lead geo-economics factors stay latent in the model for a flight route on which unsatisfactory service is offered by airlines. In order to exploit potential market demand, competition is the best policy which urges airlines to offer competitive product and services satisfying customer need.

Although the model is performing significantly very well with a relatively high R-square value, small discrepancy in prediction value may result in larger inaccuracy in passenger demand estimate because of the logarithmic functional form of the dependent variable of the regression model.

Air travel demand in Turkey has shown a twofold growth: first is traffic shift from the bus companies and second is induced demand resulting in proliferation of more frequent flyers due to more affordable ticket prices.

5.2 Conclusion

The multivariate logarithmic regression model reveals all the quantitative relationships among the used variables, which is helpful for airlines to understand the consequence of change of their decision variables or adjustment of their routing structures. Whereas, neural network performs more reliable prediction which could certainly help industry practitioners and official policy makers improve their planning and decision making.

This research was, to the date, a first attempt to model domestic air travel demand in Turkey per city pair. To decide which forecasting method to use, the general conclusion that can be drawn from this research seems to be that if values of independent variables are known or can be estimated accurately, then neural networks will give the best results. When impacts of independent variables for policy issues need to be measured, regression models are more useful.

The model efficiency may be improved for even more reliable estimation for a particular city pair, if more explanatory variables indicating bilateral interactivity between city-pairs are included in the model such as the number of call between city-pairs or credit card statistics of domestic visitors. Additionally, it is expected that having market size of alternative modes of transportation for each city pair embedded into model would most likely contribute model accuracy.

APPENDIX A Urban Population of Turkey

No	City	2010	2011	No	City	2010	2011
1	Adana	1.836.432	1.864.591	42	Kahramanmaraş	636.828	656.783
2	Adıyaman	347.236	356.595	43	Karabük	177.189	169.698
3	Afyon	365.421	370.411	44	Karaman	159.834	162.487
4	Ağrı	275.785	290.904	45	Kars	123.452	129.047
5	Aksaray	228.060	233.005	46	Kastamonu	195.059	197.704
6	Amasya	219.541	210.947	47	Kayseri	1.064.164	1.090.530
7	Ankara	4.641.256	4.762.116	48	Kilis	85.923	87.939
8	Antalya	1.392.974	1.450.209	49	Kırıkkale	233.073	233.768
9	Ardahan	33.701	37.424	50	Kırklareli	219.333	229.000
10	Artvin	89.960	91.886	51	Kırşehir	156.731	158.179
11	Aydın	588.552	599.973	52	Kocaeli	1.459.772	1.499.958
12	Balıkesir	694.926	701.213	53	Konya	1.486.653	1.527.937
13	Bartın	63.984	65.856	54	Kütahya	383.572	362.274
14	Batman	373.388	388.523	55	Malatya	480.144	498.588
15	Bayburt	37.537	40.354	56	Manisa	924.267	891.084
16	Bilecik	173.389	153.017	57	Mardin	428.899	446.226
17	Bingöl	138.069	147.081	58	Mersin	1.281.048	1.303.018
18	Bitlis	168.787	178.788	59	Muğla	350.050	362.513
19	Bolu	169.962	175.553	60	Muş	143.624	152.064
20	Burdur	159.508	152.408	61	Nevşehir	154.103	157.462
21	Bursa	2.308.574	2.359.804	62	Niğde	163.237	168.596
22	Çanakkale	269.035	268.082	63	Ordu	404.390	409.288
23	Çankırı	110.222	113.191	64	Osmaniye	346.707	354.054
24	Çorum	355.015	361.244	65	Rize	197.520	202.636
25	Denizli	641.093	655.322	66	Sakarya	646.899	664.813
26	Diyarbakır	1.090.172	1.132.351	67	Samsun	816.576	827.796
27	Düzce	194.128	198.756	68	Şanlıurfa	922.539	951.925
28	Edirne	261.920	272.294	69	Siirt	181.410	189.854
29	Elazığ	400.675	410.625	70	Sinop	107.275	109.807
30	Erzincan	134.028	125.324	71	Şırnak	269.494	290.307

31	Erzurum	489.486	505.254	72	Sivas	433.932	425.297
32	Eskişehir	681.854	700.355	73	Tekirdağ	545.481	572.359
33	Gaziantep	1.501.566	1.556.149	74	Tokat	363.944	358.872
34	Giresun	245.381	248.547	75	Trabzon	415.652	421.504
35	Gümüşhane	61.162	64.082	76	Tunceli	47.531	56.112
36	Hakkari	136.050	153.860	77	Uşak	225.570	228.785
37	Hatay	743.439	732.802	78	Van	539.619	526.725
38	Iğdır	95.550	99.550	79	Yalova	139.388	142.881
39	Isparta	311.064	277.327	80	Yozgat	268.349	269.439
40	İstanbul	13.120.596	13.483.052	81	Zonguldak	287.321	286.032
41	İzmir	3.606.326	3.623.540				

APPENDIX B Bedding Capacity of Domestic Cities

No	City	2010	2011	No	City	2010	2011
1	Adana	3.621	3.795	42	Kahramanmaraş	813	899
2	Adıyaman	489	721	43	Karabük	700	657
3	Afyon	3.652	3.300	44	Karaman	184	232
4	Ağrı	609	609	45	Kars	981	1.131
5	Aksaray	932	1.020	46	Kastamonu	884	565
6	Amasya	590	781	47	Kayseri	2.462	2.846
7	Ankara	19.231	20.936	48	Kilis	45	45
8	Antalya	306.535	331.688	49	Kırıkkale	68	68
9	Ardahan	124	174	50	Kırklareli	178	436
10	Artvin	933	933	51	Kırşehir	352	352
11	Aydın	19.483	21.497	52	Kocaeli	2.722	3.023
12	Balıkesir	8.760	9.051	53	Konya	3.496	3.658
13	Bartın	582	582	54	Kütahya	988	1.008
14	Batman	795	913	55	Malatya	1.250	1.426
15	Bayburt	-	-	56	Manisa	1.622	1.454
16	Bilecik	461	525	57	Mardin	1.098	1.682
17	Bingöl	56	56	58	Mersin	4.725	5.509
18	Bitlis	323	351	59	Muğla	88.479	90.427
19	Bolu	3.067	3.350	60	Muş	415	415
20	Burdur	173	173	61	Nevşehir	8.424	8.718
21	Bursa	6.335	8.066	62	Niğde	240	240
22	Çanakkale	4.172	4.749	63	Ordu	1.299	1.419
23	Çankırı	209	209	64	Osmaniye	120	320
24	Çorum	791	1.013	65	Rize	785	813
25	Denizli	5.592	6.044	66	Sakarya	1.569	1.661
26	Diyarbakır	1.773	1.920	67	Samsun	1.133	1.644
27	Düzce	584	739	68	Şanlıurfa	1.175	1.750
28	Edirne	1.638	1.634	69	Siirt	-	-
29	Elazığ	638	631	70	Sinop	158	314
30	Erzincan	258	517	71	Şırnak	177	177
31	Erzurum	1.953	2.031	72	Sivas	585	585

32	Eskişehir	1.899	2.261	73	Tekirdağ	1.292	1.754
33	Gaziantep	3.961	4.352	74	Tokat	501	536
34	Giresun	913	913	75	Trabzon	2.880	2.995
35	Gümüşhane	143	140	76	Tunceli	36	204
36	Hakkari	253	329	77	Uşak	418	386
37	Hatay	2.597	3.075	78	Van	973	1.140
38	Iğdır	427	422	79	Yalova	727	749
39	Isparta	1.046	1.081	80	Yozgat	426	430
40	İstanbul	62.841	68.723	81	Zonguldak	958	1.380
41	İzmir	25.688	27.831				

APPENDIX C Potential Cross City Routes

Rank	Origin	Destination	Distance	Passenger	Frequency
1	Antalya	Bursa	537	125.168	9,1
2	Bursa	Gaziantep	1.042	110.207	8,0
3	Antalya	Kocaeli	605	102.300	7,4
4	Bursa	Diyarbakır	1.281	92.740	6,7
5	Adana	Bursa	837	86.660	6,3
6	Bursa	Şanlıurfa	1.179	81.892	6,0
7	Antalya	Balıkesir	510	76.878	5,6
8	Gaziantep	Kocaeli	1.013	76.230	5,5
9	Bursa	Hatay	1.028	74.830	5,4
10	Antalya	Tekirdağ	848	70.628	5,1
11	Artvin	İstanbul	1.300	69.241	5,0
12	Diyarbakır	Kocaeli	1.252	66.638	4,8
13	Adana	Balıkesir	894	66.425	4,8
14	Balıkesir	Gaziantep	1.099	64.409	4,7
15	Adana	Kocaeli	828	63.208	4,6
16	Bursa	Muğla	541	62.788	4,6
17	İzmir	Tokat	957	61.976	4,5
18	Bursa	Erzurum	1.236	61.761	4,5
19	Antalya	Eskişehir	424	61.516	4,5
20	Kocaeli	Şanlıurfa	1.150	61.501	4,5
21	Bursa	Malatya	1.030	60.960	4,4
22	Bursa	Trabzon	1.078	60.179	4,4
23	Bursa	Kahramanmaraş	962	59.875	4,4
24	Hatay	Kocaeli	1.019	58.668	4,3
25	Bursa	Kayseri	689	57.665	4,2
26	Diyarbakır	Konya	874	57.325	4,2
27	Kocaeli	Muğla	669	57.124	4,2
28	Adana	Tekirdağ	1.071	56.841	4,1
29	Antalya	Denizli	222	55.973	4,1
30	Eskişehir	Gaziantep	893	55.718	4,0

31	Bursa	Mardin	1.367	55.432	4,0
32	Balıkesir	Kayseri	837	54.849	4,0
33	Bursa	Samsun	745	54.750	4,0
34	Bursa	Van	1.603	54.243	3,9
35	Bursa	Elazığ	1.128	54.006	3,9
36	Antalya	Tokat	876	52.713	3,8
37	İstanbul	Siirt	1.550	52.334	3,8
38	Adıyaman	Bursa	1.126	51.864	3,8
39	Gaziantep	Tekirdağ	1.256	51.727	3,8
40	Bursa	İstanbul	243	51.370	3,7
41	Ankara	Bursa	382	50.417	3,7
42	Ağrı	İzmir	1.631	50.407	3,7
43	Erzurum	Konya	934	50.261	3,7
44	Erzurum	Kocaeli	1.114	50.235	3,7
45	Diyarbakır	Eskişehir	1.132	50.042	3,6
46	Batman	Bursa	1.381	49.973	3,6
47	Kocaeli	Malatya	1.001	49.484	3,6
48	Bursa	Sivas	822	48.231	3,5
49	Balıkesir	Samsun	896	48.216	3,5
50	Kahramanmaraş	Kocaeli	933	48.131	3,5
51	Denizli	Şanlıurfa	1.110	47.680	3,5
52	Eskişehir	Şanlıurfa	1.030	47.513	3,5
53	Balıkesir	Diyarbakır	1.412	47.448	3,4
54	Konya	Mardin	886	47.207	3,4
55	Balıkesir	Şanlıurfa	1.236	47.009	3,4
56	Konya	Tekirdağ	792	46.416	3,4
57	Adana	Eskişehir	688	46.339	3,4
58	Balıkesir	Hatay	1.085	45.531	3,3
59	Kayseri	Tekirdağ	903	45.322	3,3
60	Kayseri	Kocaeli	660	45.114	3,3
61	Konya	Şanlıurfa	698	44.666	3,2
62	Elazığ	Kocaeli	1.099	44.270	3,2
63	Bursa	Nevşehir	646	44.241	3,2
64	Bursa	İzmir	322	44.016	3,2
65	İzmir	Kocaeli	450	43.902	3,2

66	Balıkesir	Muğla	392	43.895	3,2
67	Batman	Konya	970	43.724	3,2
68	Kocaeli	Mardin	1.338	43.631	3,2
69	Eskişehir	Hatay	879	43.528	3,2
70	Ankara	Artvin	980	42.938	3,1
71	Kocaeli	Van	1.527	42.576	3,1
72	Bursa	Tokat	760	42.431	3,1
73	Adıyaman	Kocaeli	1.097	42.418	3,1
74	Ağrı	Bursa	1.417	42.391	3,1
75	Balıkesir	Kahramanmaraş	1.079	42.151	3,1
76	Çanakkale	İzmir	325	42.035	3,1
77	Çanakkale	Konya	744	41.107	3,0
78	Adana	Ağrı	961	41.061	3,0
79	Konya	Samsun	640	40.687	3,0

APPENDIX D Potential New Airport

Rank	New Airport	İstanbul	Ankara	İzmir	Total Pax
1	Mersin	369.108	53.821	143.433	1.942.631
2	Aydın	243.605	95.649	33.661	1.649.325
3	Ordu	136.966	41.920	74.504	1.114.909
4	Osmaniye	110.537	30.915	59.085	780.773
5	Rize	108.620	55.904	46.214	823.199
6	Giresun	104.425	48.538	56.129	863.797
7	Şırnak	102.621	59.231	47.870	771.006
8	Bitlis	89.696	53.023	37.643	640.586
9	Aksaray	72.928	21.425	52.647	704.964
10	Çorum	69.519	22.061	63.302	880.061
11	Bingöl	68.669	39.418	37.262	475.856
12	Hakkari	65.227	45.143	31.541	605.993
13	Iğdır	63.697	41.293	26.002	508.579
14	Niğde	61.376	20.795	42.869	524.503
15	Yozgat	60.387	17.095	51.513	667.134
16	Karaman	59.043	21.736	38.511	536.585
17	Afyon	58.625	29.976	39.165	971.839
18	Manisa	57.026	40.956	15.588	1.277.106
19	Kilis	55.663	26.913	33.064	367.782
20	Gümüşhane	54.053	29.779	26.187	372.974
21	Kırşehir	53.791	14.704	43.380	513.115
22	Tunceli	51.031	31.092	23.348	340.146
23	Bolu	39.134	25.852	47.939	715.017
24	Karabük	38.712	18.632	47.109	623.782
25	Kastamonu	38.099	19.564	51.432	665.411
26	Ardahan	37.155	24.867	16.011	280.568
27	Zonguldak	36.806	25.703	55.908	859.697
28	Burdur	36.606	20.251	21.322	496.531
29	Bartın	35.786	19.323	31.873	391.449
30	Kütahya	35.391	26.440	30.543	847.694
31	Çankırı	34.687	11.019	35.570	438.608

32	Bayburt	31.771	9.678	19.037	187.077
33	Edirne	29.064	57.685	41.716	862.226
34	Bilecik	24.900	22.858	27.278	541.049
35	Düzce	24.286	20.394	35.941	636.739
36	Kırıkkale	23.813	8.907	31.806	442.313
37	Yalova	22.556	27.688	30.966	551.018
38	Sakarya	22.221	27.543	42.347	1.125.095
39	Kırklareli	21.091	44.984	32.933	725.817

APPENDIX E Airport Statistics in 2011

Ran-	**Airport**	**Domestic Passenger**	**Rank**	**Airport**	**Domestic Passenger**
1	İstanbul-Yeşilköy	13.421.536	25	Batman	173.943
2	İstanbul-Kurtköy	8.704.249	26	Çardak	168.260
3	Ankara	7.080.072	27	Ağrı	134.519
4	İzmir	6.125.076	28	Nevşehir	127.730
5	Antalya	4.516.485	29	Mardin	122.912
6	Adana	2.651.873	30	Kahramanmaraş	95.740
7	Trabzon	2.190.503	31	Bursa Yenişehir	67.410
8	Diyarbakır	1.714.423	32	Çanakkale	60.543
9	Muğla-Bodrum	1.396.493	33	Sinop	58.438
10	Gaziantep	1.170.025	34	Balıkesir Körfez	56.933
11	Samsun	1.064.301	35	Merzifon	48.035
12	Van	1.055.358	36	Adıyaman	45.346
13	Kayseri	968.942	37	Tekirdağ Çorlu	42.839
14	Erzurum	788.128	38	Siirt	31.420
15	Muğla-Dalaman	696.644	39	Tokat	30.516
16	Hatay	553.527	40	Isparta	20.724
17	Malatya	553.142	41	Uşak	15.267
18	Konya	530.509	42	Eskişehir	12.508
19	Elazığ	513.804	43	Kocaeli	11.851
20	Kars	377.584	44	Balıkesir Merkez	6.674
21	Şanlıurfa	224.677	45	Antalya Gazipaşa	4.192
22	Sivas	221.049	46	Gökçeada	1.106
23	Erzincan	207.074	47	Zonguldak	160
24	Muş	195.784			

Source: State Airport Administration

REFERENCES

Abdullah O. Ba-Fail and Seraj Y. Abed (2000). The Determinants of Domestic Air Travel Demand in the Kingdom of Saudi Arabia. *Journal of Air Transportation World Wide*, *5*, 72-86.

Abdullah Omer Ba-Fail (2004). Applying Data Mining Techniques To Forecast Number of Airline Passengers In Saudi Arabia (Domestic and International Travels). *Journal of Air Transportation*, *9*, 100-116.

Adakunle J. Aderamo (2010). Demand for Air Transport in Nigeria. *Journal of Economics*, *1*, 23-31.

C. Lewis (1982). Industrial and Business Forecasting Methods. Butterworth Scientific, London.

Chaug-Ing Hsu and Su-Miao Liu (2003). Predicting City-Pair Air Passenger Traffic Using Grey Topological Forecasting Model. *Journal of the Eastern Asia Society for Transportation Studies, 5*.

C. Wu, J. Han, and Y. Hayashi (2011). Airport attractiveness analysis through a gravity model: A case study of Chubu International Airport in Japan. *Eastern Asia Society for Transportation Studies, 8,* CD-ROM (100350).

Dail Umamil Asri and Yoriyasu Sugie (2003). Simultaneous Demand Model for Passenger Travel. *Proceedings of the Eastern Asia Society for Transportation Studies, 4.*

Dipasis Bhadra (2003). Demand for Air Travel in the United States: Bottom-up Econometric Estimation and Implications for Forecasts by Origin and Destination Pairs. *Journal of Air Transportation*, *8*, 19-56.

Dougherty, M. (1995). A Review of Neural Networks Applied to Transport. *Transportation Research Part C*, *3*, 247-260.

D. U. Asri and Y. Sugie (2003). Simultaneous demand model for passenger travel: A case study of Indonesia. *Eastern Asia Society for Transportation Studies, 4*, 869-884.

F. Kopsch (2012). A demand model for domestic air travel in Sweden. *Journal of Air Transport Management*, *20*, 46-48.

Hruschka, H. (1993). Determining market response function by neural network modeling: A comparison to econometric techniques. *European Journal of Operational Research*, *66*, 27-35.

Hsu, C. I. and Y. H. Wen (1998). Improved grey prediction models for the trans-Pacific air passenger market. *Transportation Planning and Technology*, *22*, 87-107.

J.J. Montano, A. Palmer (2003). Numeric Sensitivity Analysis Applied to Feedforward Neural Networks. *Neural Computer & Application, 12*, 119-125.

Jorge Calderon, J.D. (1997). A demand model for scheduled airline services on international European routes. *Journal of Air Transport Management*, 3, 23-35.

Joyce Dargay and Mark Hanly (2001). The Determinants of the Demand for International Air Travel to and from UK, ESRC Transport Studies Unit, *Centre for Transport Studies,* University College London, 59, 1-14.

J.Scott Armstrong and Michael C. Grohman (1972). A Comparative Study of Methods for Long-Range Market Forecasting. *Management Science*, *19*(2), 211-221.

Kuo, S.Y., L. C. Shiau and Y. P. Chang (2010). Air transport demand forecasting in routes network by artificial neural networks. *Journal of Aeronautics Astronautics and Aviation Series B*, *42*(1), 67-72.

K. W. Kim, H. Y. Seo, and Y. Kim (2003). Forecast of domestic air travel demand change by opening the high speed rail. *KSCE Journal of Civil Engineering, 7*, 603-609.

Lawrence R. Weatherford, Travis W. Gentry and Bogdan Wilamowski (2003). Neural network forecasting for airlines: A comparative analysis. *Journal of Revenue and Pricing Management*, *1*(4), 319-331.

Makridakis, S., Wheelwright, S. C., (1989). Forecasting Methods for Management. John Wiley & Sons, New York.

Martijn Brons, Eric Pels, Peter Nijkamp and Piet Rietveld (2002). Price Elasticities of Demand for Passenger Air Travel. *Journal of Air Transport Management*, *8*, 165-175.

Matthew G. Karlaftis (2008). Demand Forecasting in Regional Airports. Straer 7, 100-111, Tr. 312.

Md. Jobair Bin Alem and Dewan Masud Karim (1998). Air Travel Demand Model For Domestic Air Transportation in Bangladesh. *Journal of Civil Engineering*, *26*, 1-13.

Proctor, R. A. (1992). Marketing decision support systems: A role for neural networks. *Marketing Intelligence and Planning*, *10*, 1, 22.

R. A., Ippolito (1981). Estimating airline demand with quality of service variables. *Journal of Transport Economics and Policy*, *15*, 7-15.

Richard T. Carson, Tolga Cenesizoğlu and Roger Parker (2011). Aggregate Demand for USA Commercial Air Travel. *International Journal of Forecasting*, *27*, 923-941.

Rob Law, Norman Au (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management*, *20*, 89-97.

Senanu Ashiabor, Hojong Baik and Antonio Trani (2007). Logit Models For Forecasting Nationwide Intercity Travel Demand in the USA. *Journal of the Transportation Research Board*, Washington, D.C., 1-12.

Seraj Y. Abed, Abdullah O. Ba-Fail and Sajjad M. Jasimuddin (2001). An Econometric Analysis of International Air Travel Demand in Saudi Arabia. *Journal of Air Transport Management*, *7*, 143-148.

S.C. Wirasinghe and A.S. Kumarage (1998). An Aggregate Demand Model for Intercity Passenger Travel in Sri Lanka. *Transportation*, *25*, 77-98.

Tobias Grosche, Franz Rothlauf and Armin Heinzl (2007). Gravity Models for Airline Passenger Volume Estimation. *Journal of Air Transport Management*, *13*, 175-183

Wenbin Wei and Mark Hansen (2006). An Aggregate Demand Model for Air Passenger Traffic in the Hub and Spoke Network. *Transportation Research Part A*, *40*, 841–851.

Zhang, X. (1994). Time Series Analysis and Prediction by Neural Networks. *Optimization Methods and Software, 4*, 151-170.