

Forecasting seat sales in passenger airlines:  
introducing the round-trip model

by

Mehrdad Varedi

A thesis  
presented to the University of Waterloo  
in fulfilment of the  
thesis requirement for the degree of  
Master of Applied Science  
in  
Management Sciences

Waterloo, Ontario, Canada, 2010

© Mehrdad Varedi 2010

## **Author's declaration**

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## **Abstract**

This thesis aims to improve sales forecasting in the context of passenger airlines. We study two important issues that could potentially improve forecasting accuracy: day-to-day price change rather than price itself, and linking flights that are likely to be considered as pairs for a round trip by passengers; we refer to the latter as the Round-Trip Model (RTM). We find that price change is a significant variable regardless of days remaining to flight in the last three weeks to flight departure, which opens the possibility of planning for revenue maximizing price change patterns. We also find that the RTM can improve the precision of the forecasting models, and provide an improved pricing strategy for planners.

In the study of the effect of price change on sales, analysis of variance is applied; finite regression mixture models were tested to identify linked traffic in the two directions and the linked flights on a route in reverse directions; adaptive neuro-fuzzy inference system (ANFIS) is applied to develop comparative models for studying sales effect between price and price change, and one-way versus round-trip models. The price change model demonstrated more robust results with comparable estimation errors, and the concept model for the round-trip with only one linked flight reduced estimation error by 5%. This empirical study is performed on a database with 22,900 flights which was obtained from a major North American passenger airline.

## **Acknowledgement**

I would like to have my sincere appreciation from Professor Benny Mantin, who kindly accepted to supervise me and introduced me to the rich area of revenue management. I will always cherish his encouragement, keen support and valuable guidance.

I would like to appreciate the kind support of Professor Juan Vera that always had his door open for my questions during this study.

I would also like to thank Professor Jeff McGill from Queen's school of business, Kingston, ON, Canada, for his kind reply and his clarifying notes on my questions.

Many thanks to my thesis readers: Dr. Mark Hancock and Dr. Qi-Ming He, for their careful reading, support, and their valuable feedback.

## **Dedication**

To my family:

my wife, Fariba, my brothers, Behzad and Mehran, and my parents, Nahid and Hossein

who are love of my life.

## Table of contents

Author’s declaration.....	ii
Abstract .....	iii
Acknowledgement .....	iv
Dedication.....	v
Table of contents .....	vi
List of tables .....	viii
List of figures.....	ix
1. Introduction.....	1
2. Literature review .....	2
2.1. Dynamic Pricing.....	3
2.2. Forecasting .....	4
2.3. Mixture Models .....	7
3. Data .....	8
3.1. Source and data specificities .....	8
3.2. Data structure .....	8
3.3. Basic patterns in data.....	9
3.4. Important independent variables .....	14
4. Price change effect on sales .....	15
4.1. Price change characteristics .....	16
4.2. Analysis .....	18
4.3. Conclusion .....	21
5. Roundtrip models .....	22
5.1. Linked days .....	23
5.2. Linked flights .....	30
6. Fuzzy forecasting .....	34
6.1. Advantages of using a fuzzy model .....	34
6.2. Comparing the effect of price and price change in sales forecasting.....	36
6.3. Comparing the efficiency of one-way and round-trip models .....	38
6.4. Results.....	40
7. Conclusion and further research.....	41

References .....	43
Appendix I - Error Measures .....	46
Appendix II - Linear models for connected days (subsection 5.1).....	48
Appendix III – Selection criteria for the number of member functions in a fuzzy system.....	51

## List of tables

Table 1: ANOVA results to test significance of price change and days to flight as main factors on sales	19
Table 2: Homogenous subsets in days remaining to flight departure in terms of average values of sold seats in each subset .....	20
Table 3: Homogenous subsets in price change classes in terms of average sales .....	21
Table 4: Analysis for significance of coefficient of each variable in two customer segments .....	33
Table 5: Comparing errors (RMSE) for training and checking datasets in price and price-change model	36
Table 6: Comparing errors (RMSE) for training and checking datasets in price and price-change model for two consequent price change and three prices.....	37
Table 7: List of variables and the relative number of member functions in one-way and round-trip models .....	39
Table 8: Comparative table for the one-way and the round-trip models' errors .....	39

## List of figures

Figure 1: Average price (left vertical axis) and average sold seats (right vertical axis) in different days remaining to flight for two directions of the three routes .....	10
Figure 2: Average sold seats as a result of price change and in its absence are compared; the average absolute value of price change trend in the last 5 weeks to flight departure is demonstrated .....	13
Figure 3: Price change distribution in last three weeks before a flight departure .....	17
Figure 4: Enplanements in two directions in a randomly selected month for the three routes .....	24
Figure 5: BIC graphs for selecting number of segments in each route .....	27
Figure 6: The traffic in 17 days on the reverse direction as independent variables are used to estimate a target day traffic; the coefficients of each in the resulting regression model are drawn; days with significant effect have filled bars. ....	29
Figure 7: BIC values for one, two and three segments in the returning enplanements .....	31
Figure 8: Rootograms for the two segments identified in return flight choice .....	32

## 1. Introduction

In many businesses situations, firms seek to maximize revenue by selling a limited amount of a perishable product. For example, in the passenger airplanes, the number of seats is fixed, and they can be sold only until the departure time. Given the fixed inventory of seats, airlines forecast demand in order to set prices accordingly. Forecasting demand in passenger airlines is a challenging task, because the knowledge about the customers' preferences is limited, and the business environment rapidly changes.

There are different approaches to improve forecasting. One approach is focused on understanding the customers' decision-making processes; another approach focuses on mathematical tools and models (i.e. while in the customer oriented approach, the effect of weekdays on sales is studied, in the other one, the neural network models are compared with the regression models). One of the customers' behaviour oriented models for forecasting, which made a change in revenue management is customer segmentation based on their arrival time (i.e., how far in advance they purchase their tickets), and relating it to their willingness to pay. Using this basic concept regarding customers' behaviour, in 1985 American airlines offered the Super Saving program that helped it to compete with PEOPLExpress after the deregulation of the airline industry in the U.S.A.

We want to improve sales forecasting with a customer behaviour orientation. According to Webber (2004) most of travelers have round trips with an average staying period of less than four days in the local travels in Canada. We think that only a few roundtrips exist that travelers prefer to select and their preferences is not spread around all the options evenly. This creates new possibilities for improved pricing mechanisms, by managing price pairs instead of planning for prices of each flight separately.

Another idea for a better forecasting is more attention to the travelers' reactions to price change. The customers look at the airline prices as a long term increasing trend with limited local fluctuations and they should decide for buying a ticket with a "good price" in the absence of more accurate information about future prices.

According to our observation in our data from a major airline, the average seat sales in price change situations is approximately 25% - 30% higher than days with no change in price in the last five weeks before flight (see Figure2), so one of the studies is identifying the effect of price change on the seat sales.

The scope of this research is revenue management in the passenger airlines and the results can be applied for similar modes of transportation. This research is price based Revenue Management (RM) rather than capacity based RM. We use analysis of variance to study price change effect on the seat sales, employ mixture models to identify customer segments when finding a relation between aggregate passenger traffic in both direction of a route. Finally we use fuzzy systems to develop forecasting models to make comparison between the selected factors which affect the seat sales.

The study starts with a literature review, continues with an explanation about available data in section 3, studying price change effect on sales in section 4, the introducing and study of round trip models in section 5, fuzzy forecasting models in section 6, and in section7, the conclusion and possibilities for future work is discussed.

## **2. Literature review**

The focus of this research is on improvement of the sales forecasts which is an important component of revenue management in the light of rapid changes in the market demand. On-line sales channels, rapid changes in economical situations, as well as unpredictable events such as epidemic disease or terrorist attacks make the long term sales forecasting models less reliable for everyday operations, as a result, identifying core factors and constant behaviour in customers that affect their decisions for buying tickets become important.

The three main issues in this research is studying the significance of price change effect on sales, which is more close with dynamic pricing, mixture models and their general applications , as well as their

specific use in revenue management, and finally mathematical methods that have been used for deterministic sales forecasting with special focus on fuzzy systems.

## **2.1. Dynamic Pricing**

One of the interesting findings in this research is increase in the average sales, as a result of price change (either upwards or downwards). This finding is related to similar literature on revenue management and dynamic pricing.

Dynamic pricing is price-based Revenue Management (RM), which is formerly employed in passenger airline industry after deregulation; it is the main RM policy in Low Cost Carriers (LCCs) and where there is only one class in seats (Talluri and van Ryzin (2005)). Talluri and van Ryzin explain different sectors that make benefit of dynamic pricing such as retails, manufacturing, and e-business; mark-down pricing, discount airline pricing, and consumer-packaged goods promotions as some examples in dynamic pricing are discussed in their book and dynamic pricing optimization problem for deterministic and stochastic demand is formulated.

Gallego and van Ryzin (1997) argue that when there is an option to select between price-based and quantity-based RM, the price-based RM is preferred. The reason they mention is that quantity-based RM reduces sales by rationing but when there is price flexibility, the demand can be controlled smoothly by changing the price. This point of view demonstrates importance of this approach in revenue management and the efforts in this direction.

Some efforts to find revenue maximizing patterns by price changes can be found in Feng and Xiao (2000); they have considered a set of prices, assuming the demand to have a Poisson distribution, and demonstrated that concave subsets in the price set, maximize revenue. Similar research can be found in Bitran and Caldentey (2002), that have provided an overview on this kind of studies and they have also proposed a model. Having a pattern for maximizing revenue is a wonderful idea although frequency of

price change in these patterns or volatility in prices may have an important effect on sales, a subject which is rarely studied.

Gillen and Mantin (2009), have studied price volatility and introduced a measure for it. They have demonstrated that in two weeks prior to flight, volatility increases significantly and it does not have any relation with market structure although price volatility depends on type of carriers (LCC or legacy carrier). This finding was the leading idea that price change may have some effect on sales and our data analysis was inspired from this idea that not only the shape of a price set might affect the revenue but also frequency of these changes may affect sales as getting closer to flight departure.

After this short review of the literature in dynamic pricing, it seems that finding any revenue maximizing pattern in prices is difficult specially when considering the competition in the market, although the increase of price volatility is worthy to be studied further. As a common judgement, in dynamic pricing, a price increase is usually used to increase marginal revenue, while a demand reduction is expected, although we have found that it might be slightly different and people may even buy more when price increases within a certain range. The study of this aspect of travelers' reaction may explain higher volatility in prices in the last days before flight departure and it might be a source for higher revenue generation.

## **2.2. Forecasting**

In dynamic pricing, price based RM is a continuous process of forecasting and optimization, although recently some researchers are offering new adaptive approaches which does not need forecasting. This approach might be efficient although they may still need further research to become a perfect replacement for the existing model. Some of these studies are mentioned here to be considered by interested people. See Levina, et al. (2009) and Farias (2007), however in this research we intent to improve the forecasting in price-based RM.

Before reviewing research in forecasting, it is important to separate demand from sales. Censored data is a source of errors in demand forecasting, because the demand for seats is not recorded when seats are sold out; therefore, historical booking data reflect the sales and not the demand. The problem becomes more complicated when a flight has different seat classes, and scarcity or price of one class affects the demand of others. McGill (1995) studies this problem, i.e., analysis of multiclass demand subject to capacity constraint using censored data. Weatherford and Pölt (2002) provided another solution to identify the real demand level. In our study, we rarely experience data censorship due to stock-outs (i.e. when all seats on the plane were sold).

Overbooking and forecasting studies are somehow related and a complete overview of research can be found in McGill and van Ryzin (1999) although we will mention a few of them and add some additional research, which are among the outstanding efforts and are more related with the scope of this study.

Littlewood (2005) which is originally published in 1972 made one of the first contributions to the field and used exponential smoothing for forecasting. Sa (1987) compared regression results with ARIMA and historical averages and concluded that, regression models provide higher accuracy. His model did not have price or price-related factors, and sales history was the main factor for his forecast. He had historical averages of reservations, average booking in the last five flights for each day to flight, weekdays and week of the year in his regression model.

One of the most outstanding jobs on probabilistic demand models is carried out by Lee (1990) in his Ph.D. thesis. In addition to his comprehensive approach to formulate the probabilistic demand models, he has done a simulation which indicates that each 10% improvement in accuracy of forecasting sales increases revenue by 1-5%. This result was confirmed again in the study of Weatherford and Belobaba (2002) with a slight difference.

Taneja (1978) proposed quantitative, qualitative, and decision analysis approaches for forecasting; he has provided comprehensive explanations for the development and appraisal of regression models with

applications in forecasting and planning air transportation services. In his book, he has included researchers' work in deterministic modeling for air passenger demand models.

Neural Network (NN) models have been used to forecast passenger traffic by Nam and Schaefer (1995), they concluded that NN models have better performance in forecasts, comparing with conventional methods in the airline industry such as regression or time series. Similar research is also done by Weatherford, Gentry, and Wilamowski (2003), comparing moving average, exponential smoothing, and regression models with the simplest NN models, which demonstrated a better performance of a NN model. They used the mean absolute percentage error (MAPE) for error measurement and took separate datasets for training, and testing to evaluate the models' performances.

The trace of fuzzy systems' application in airlines can be found in the decision support system for sales forecasting offered by Kuo and Xue (1998), evaluating the effectiveness of promotions. The authors used a fuzzy-Delphi<sup>1</sup> method to collect their fuzzy input; Kuo, in European journal of operations research on 2001, offered the idea of using genetic algorithm to find initial weights for fuzzy neural network, and concluded the superior performance of fuzzy neural network model over simple neural network model.

Becher (2008) has implemented a revenue management system with a fuzzy approach. His work covers the concept and theory of fuzzy systems with complete explanation of its components. Becher has used the rule based setting of fuzzy systems effectively to cover management expectations from a decision support system for revenue management, providing a flexible framework to include any further assumption, policy or criteria; he also offers methods to customize the proposed fuzzy system, using adaptive methods. In this research, Becher starts with analysis and development of capacity and price models and proposes a mixed model at the end.

---

<sup>1</sup> The Delphi method is a systematic, interactive forecasting method which relies on a panel of experts.

As a result of reviewing literature in forecasting section, fuzzy is selected as the suitable method to implement our forecasting models; this research aims to cover a gap in forecasting by proposing additional variables after segmenting travelers and identifying linked flights, using mixture models.

### **2.3. Mixture Models**

A single rule may not be sufficient to explain consumers' decisions; decisions of each person in a group might be taken under the effect of different goals and their personal criteria. In such situations, sometimes identifying segments within the group may result in more accurate models to explain peoples' choices. So instead of proposing a single relation between independent and dependent variables, one relation for each segment can be proposed; such models are called mixture models. Mixture models have found applications in a wide range such as: biology, medicine, physics, economics and marketing; an example of mixture models in market segmentation, is Kamakura and Wedel (2001), replacing traditional cluster analysis and regression methods with it.

Finite mixture models have been extended in the 1990s by mixing standard and generalized linear models, DeSarbo and Wedel (1995); they develop a mixture model approach to simultaneously estimate the posterior membership probabilities of observations to a number of latent classes; they also use EM algorithm to estimate their model parameters. A comprehensive review about the existing application and theory of finite mixture models can be found in McLachlan and Peel (2000).

Leisch (2004) also mentions some applications of mixed modeling and proposes a general framework to fit discrete mixture of regression models; he investigates different Expectation Maximization (EM) algorithms for this purpose which are usually used for identifying optimum number of segments and avoids overfitting by proposing suitable number of variables in a model.

In passenger airline research, mixture models are also applied by Koenigsberg, Muller and Vilcassim (2008) to explain arrival of travelers in two segments using a Poisson latent class model; their main

objective in this research is evaluating the benefit of last minute deals, trying to compare different situations by segmenting high valuation and low valuation customers, in base of their arrival time.

### **3. Data**

In this section, we plan to explain the available data which is used for analysis and hypothesis testing throughout this research. In order to do that, source of data and some specific information about it, such as structure of data tables, and number of records in each table is mentioned. Some basic patterns in the data is demonstrated in two series of graphs. In the first group of graphs, average price and average number of sold seats in the last 30 days remaining to flight are simultaneously demonstrated, which may convey some information for professionals about the type of markets that the datasets are selected from. In the second set of graphs, the average sold seats when price is constant and when it is changing in the last five weeks to flight departure are demonstrated, which helps to understand the reason for proposed hypothesis in section 4.

#### **3.1. Source and data specificities**

The data for this research is obtained from a major North American airline, that we are not authorized to mention its name, because of the company's policy. This data set contains the sold seats information on a daily basis for all flights from three non-stop routes in both directions. All together, the dataset contains over 500,000 sales records from 22,900 flights over 20 months.

#### **3.2. Data structure**

The firm has provided us with daily sales data for three routes in both directions. The information are organized in two tables. One of them has general information about a flight which is mentioned under "General flight information" and related sales during the sales period for each flight are reported in the other table, "Sales detail information". The fields in each table are as follow:

- General flight information

- Flight direction containing a pair of three characters for the origin and destination
- Flight number
- Flight departure date
- Time of flight (hour: minute)
- Airplane capacity
- Sales detail information
  - Snapshot-date in which (recorded only on days that transaction occurred)
  - Total bookings per day
  - Number of cancelations
  - A subset of total bookings for regular customers
  - Total transacted value (i.e., total daily revenue), also missing snapshot dates were recorded as having the last observed price

### **3.3. Basic patterns in data**

As it was mentioned in the beginning of this section, some graphs are provided to give information and create some questions for readers that will be answered in the next sections. As it is pointed out in subsection 3.1, our sales data are from three routes in their both directions. On these three routes we have six cities that are called A, B, C, D, E, and F. Each direction on a route is named with a pair of the city names, i.e., AB stands for the flight direction, from origin A into destination B. In Figure 1, average price and average number of sold seat are demonstrated during the last four weeks remaining to flight departure on each route. There are six graphs in three rows, each of the two graphs on a row demonstrate trends in two directions between two cities.

Difference in demand and average price in the two directions of each route can be observed. Also the trend of the average number of seat sold is increasing in the last ten days remaining to flight in all routes.

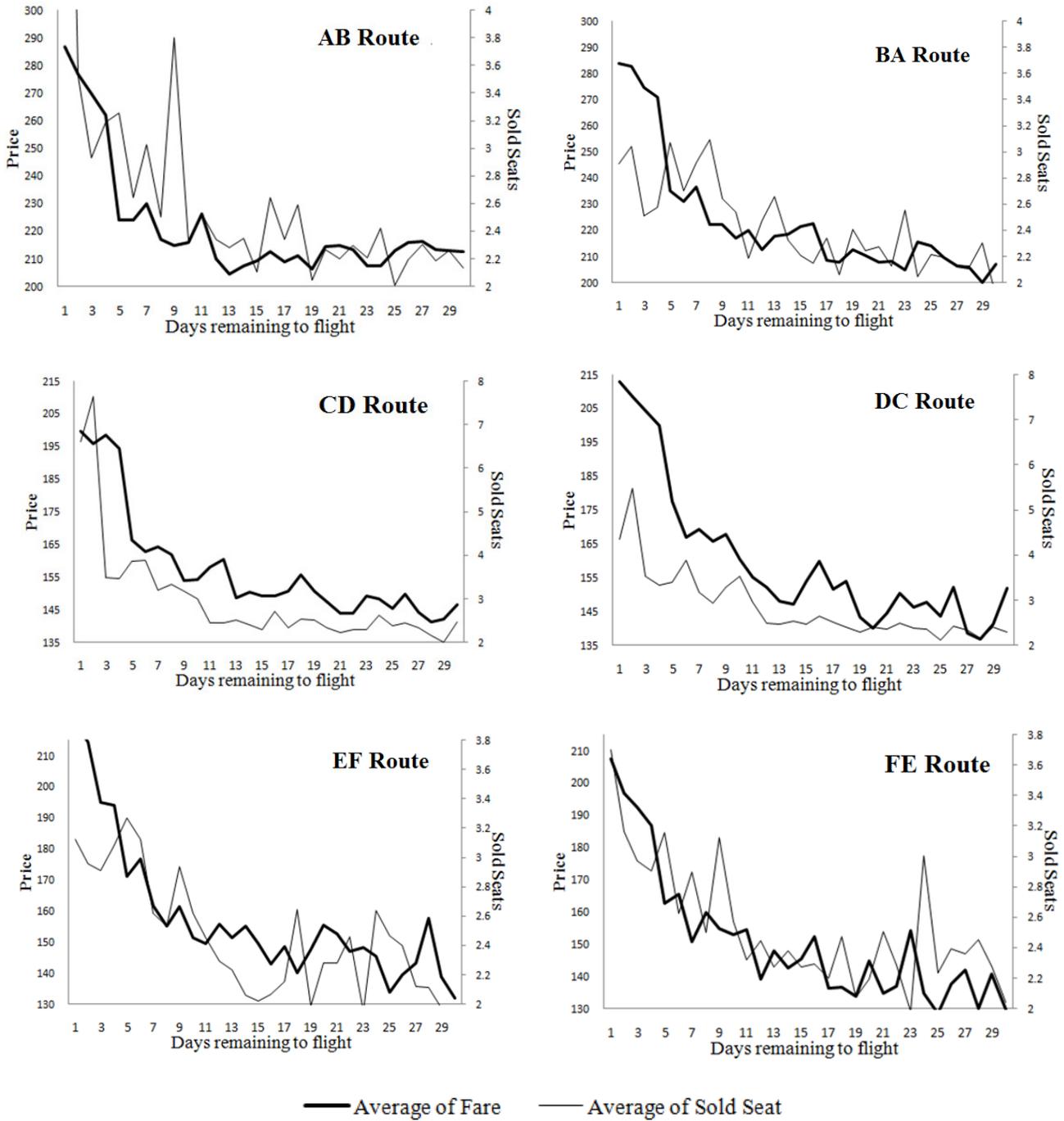


Figure 1: Average price (left vertical axis) and average sold seats (right vertical axis) in different days remaining to flight for two directions of the three routes

In Figure 2, the goal is demonstrating the higher average of seat sold, when price changes comparing with the time that price remains constant. Each graph demonstrates the trend on a route in both

directions, because the goal is comparing the three routes rather than two direction of the same route, which indicates higher sales when price changes.

In the title of the graphs in Figure 2, the same notation for cities and flight directions is used. The data which is demonstrated with an empty bar stands for the average number of seat sold when price changes and the line shows the relative average of the price change on each day. The grey bars show the average number of seats sold, when price remains constant in a certain day. Certainly, no line in the graph is attached to these data, because the relative average of price change is always zero. The data are taken for the days in the last seven weeks remaining to flight departure.

In Figure 2, the average number of seat sold in price change situations is approximately 25% - 30% higher than seats sold, when price remains fixed. Realizing Gillen and Mantin (2009), that observed significant increase in price volatility in the last two weeks to flight departure, and considering the increase in the average number of seats sold brings the analysis later in subsection 4.2., which investigates the connection between price change and the average number of seat sold. Better understanding of this relation, may help to improve revenue generation not only in the last days remaining to flight but throughout the selling period. Another point in Figure 2 is the sporadic price change patterns and no specific shape or repeating set of prices.

As it was mentioned in the literature review, Feng and Xiao (2000) proposed concave subsets in the price set to maximize revenue, they made this inference by comparing a group of price change patterns. In our data, concave price changes can be observed although a visual assessment demonstrates that sales is more influenced by price change rather than a specific pattern in the data<sup>2</sup>. Also it is unlikely that a predictable pricing policy (i.e. a concave pattern) can be applied as a successful pricing policy at the existence of market competition, and strategic<sup>3</sup> customers.

---

<sup>2</sup> For a better comparison, it requires further study and this interpretation is based on a visual assessment

<sup>3</sup> Strategic customers are the customers that optimize their own purchase behaviour in response to the pricing strategies of the firms (see Talluri and Ryzin (2005), section 5, page 182).

Bitran and Caldentey (2002) overview dynamic pricing models; they conclude that demand learning and demand substitution effects are open subjects that can improve dynamic pricing; they have proposed more research to better learn the demand in each market.

As a conclusion, proposing a certain pattern as a revenue maximization solution might not be applicable and ad-hoc decisions should be made based on factors that affect demand such as the rival price positioning and previous responses of the market to price or price changes. The information to deal with the competition effect is not available, but it is still possible to study the link between demand and the price changes, however it is expected that this would affect the precision of this study. This enables passenger airlines to decide about the next price change direction and its size to achieve the expected sales.

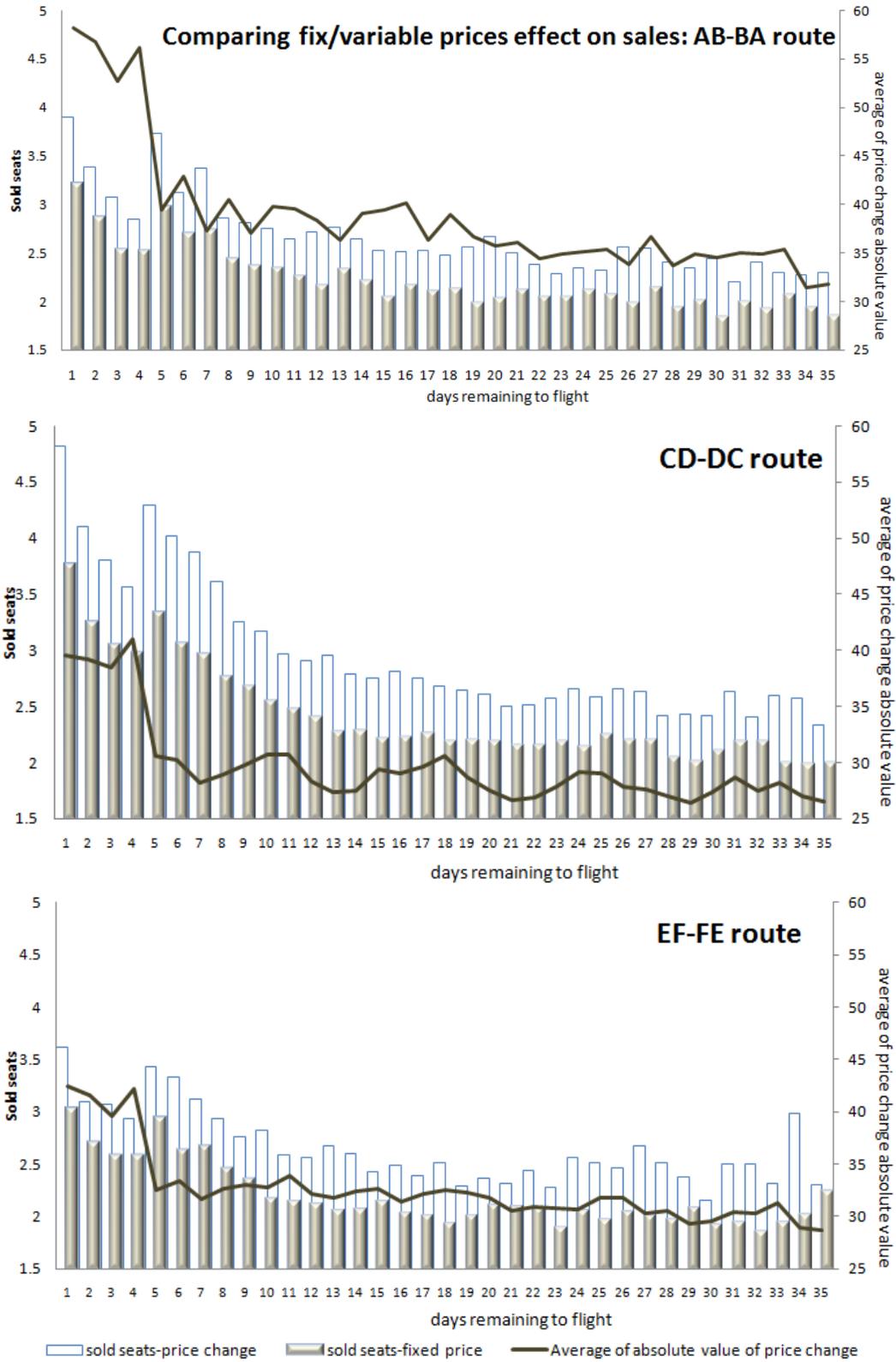


Figure 2: Average sold seats as a result of price change and in its absence are compared; the average absolute value of price change trend in the last 5 weeks to flight departure is demonstrated

### **3.4. Important independent variables**

This section intends to identify the potentially important independent variables which will be further investigated in the next sections or will be used to propose a forecasting model. One of the approaches to identify such independent variables is analysis of the processes that somehow affect the target variable and observing simultaneous trends of dependent and independent variables.

As an example, a relation between the average number of seat sold and the time remaining to flight or weekdays can be observed in Figure 1, these variables are considered by the other researchers such as Sa (1987); he developed some ARIMA models to estimate the demand using the time remaining to flight departure, the sequence of the previous demands, and weekdays. Based on the observations from Figure2, it is plausible to propose price change or price as important factors in forecasting the average number of seat sold.

In the process of purchasing a ticket, there are usually some options available in online shops, which enable customers to find the best fare, on the day before or after the selected date, so this difference can be considered as an important variable.

So some of the potential independent variables can be summarized in the following list:

- Days remaining to departure
- Sequence of 'n' previous sold seats – where 'n' is a positive integer
- Sequence of 'n' previous prices (starting from today) – where 'n' is a positive integer
- Sequence of 'n' previous price changes– where 'n' is a positive integer
- Price difference between the flight for which demand is predicted and related flights (on the same day, one day before or after)
- Seasonal factors for the time that a sales happens
- Seasonal factors for departure date (month, weekday)

In order to keep forecasting models small and efficient, it is essential to identify most important variables and identify a priority list, however some variables such as price or price change and time remaining to flight should be always a part of any forecasting model, because in optimization step of dynamic pricing, such relations are required (for more information about a deterministic dynamic pricing model, see section 7).

There are other important issues in the analysis of the ticket purchasing process, which require more efforts to include them as a variable in forecasting sales, i.e., it is a common behaviour to buy tickets to a destination and returning tickets to the origin in most of our travels, this means that if price or price related factors would be important for a traveler, then these variables should be minimum in a round trip flight and not just in one direction. This might be an important point to be considered, because it can define the price competition in form of a series of linked flights, an idea which is not mentioned in any literature to the author's knowledge to date, although there are some problems to connect one flight with a finite set of returning flights. In contrast, it is not unexpected that the returning flights would be so numerous, that considering their effect wouldn't be significant. Both mentioned ideas might be plausible speculations however the way to figure out the importance of returning flights and the way to identify them will be explained in section 5 and 6.

#### **4. Price change effect on sales**

In section 3, a visual assessment of Figure 2 demonstrated 25% - 30% higher seat sales on the days with a price change, than the days without a change in price. This pattern was demonstrated for the last five weeks before flight departure.

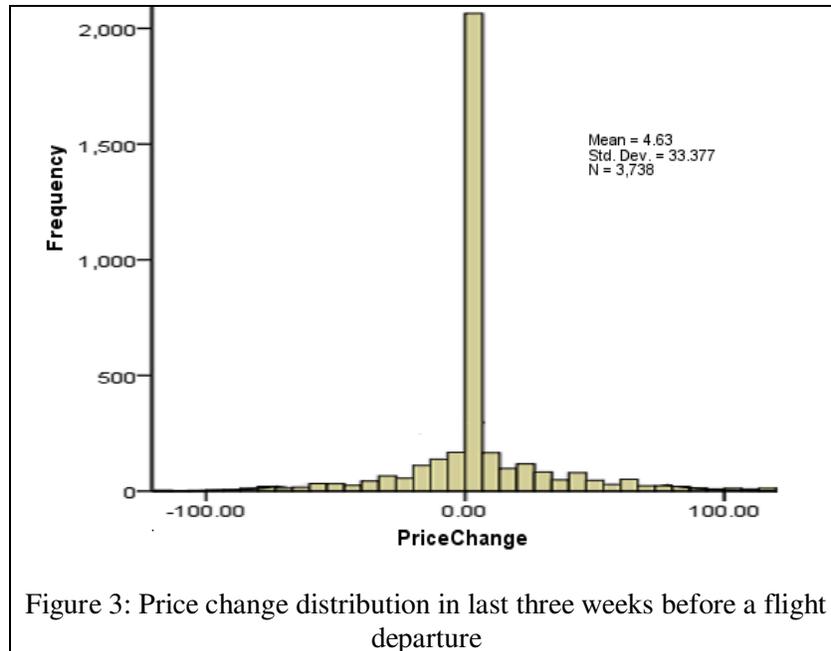
In this section we demonstrate that this difference is statistically significant. For this purpose, the frequency of price changes for different values is provided in a histogram to show the frequency and size of price change and comparing the no-price change situations with price change occurrences. Afterwards, in subsection 4.2, an analysis of variance (ANOVA) is carried out to show the difference

statistically. At the end, in the conclusion section, the results of the analysis are summarized, and building a model to compare sales forecast precision with price change and price; this comparison is done in section 6.

#### **4.1. Price change characteristics**

In this section, we aim to identify the relation between price change and sales and this requires having an idea about price change behavior in the dataset. Understanding more about size and direction of price change and its distribution may give us good information. This information can lead us to understand whether the price change is a rare event or it happens quite frequently. It is also good to know the reason for price change and ask questions about seemingly odd behavior in its size and distribution.

There were 3737 observations in this analysis, and the data were captured from the last three weeks to flight departure, for an evening flight. In Figure 3, the frequency and size of price change is demonstrated in a histogram. There are 30 different segments between -100 and 100, which is the considered range for price change; as a result of this number of segments, there might be a maximum price difference of \$6.67 between the values of the grouped price changes in one segment. Additional data which are provided in Figure 3 indicates that the average of price change is 4.63 with an standard deviation of \$33. These price change values are from a flight with an average price of around \$175.



The long bar in the middle of Figure 3, is an indicator of a large number of values in the range of 0 to 6.67, although a large number of them are expected to indicate no price change; the number of constant price occurrences is higher than any other price change segment, although if this number is compared with the cumulative number of price change occurrences, regardless of their size or sign, the probability of having a price change or a fixed price are almost the same in the selected period.

When studying price changes, very large price decreases, especially in the last days before a flight are observed; the reason might be a large number of unsold seats or might be the result of competition in the market. Price increase also happens as the departure date gets closer, which is a well known phenomenon for passengers. To better understand the reasons of price changes, Mantin and Koo (2009) have done a research which relates the dynamic price dispersion with demand characteristics and provide a clear explanation about price change and factors that intensify it.

In addition to existence of large positive or negative price change, an almost even distribution of positive and negative price change can be observed in Figure 3; considering the low occupancy level, around 60% in flights, makes it hard to explain the existence of so many occurrences of price increases, which

sometimes are around 40% of the price. By simultaneous consideration of the symmetrical distribution of price change and by remembering the positive effect of price change on sales in Figure 2, it seems unclear that how exactly the effect of price change on sales can be explained. Are there higher sales when only price decreases, and price increases are for making the possibility of more decreasing prices to attract customers or the positive price change situations may also achieve higher sales?

As a summary of the above discussions, better understanding of the price change effect on sales may help to make benefit of these situations to manage revenue more efficiently; In subsection 4.2, this relation is investigated by carrying out an analysis of variance on the available data.

## 4.2. Analysis

To identify the effect of price change on sales, an analysis of variance (ANOVA<sup>4</sup>) was carried out with *price change* and *days remaining to flight* as independent factors, and the number of seats sold as the dependant variable. For this purpose, a daily flight in the AB-BA route is selected; this particular flight had the largest dataset and the data is taken from 178 flights in their last three weeks across 18 month.

As Figure 3 demonstrates, price changes are spread around a wide range, and rarely repeating price changes could be found, so price change couldn't be entered as a factor in ANOVA. To overcome this issue we define classes for the price change values and enter them into the ANOVA as nominal factors; seven classes were set, also these class ranges were identified as percentage of the average price, to make this analysis comparable in different route and different prices; low, medium and high classes, for price increase were: (0%, 5%], (5%, 15%], and (>15%); similarly, high, medium and low price decrease classes were: (<-15%), [-15%,-5%), and [-5% , 0%); in addition, one class for no price change was considered separately.

The reason for selecting seven categories of price change is an initial guess about the number of distinct categories in terms of their effect on sales (and actually an excessive number of categories) which will be analysed using a post-hoc test later and the homogeneous categories will be identified. I.e. the post-

---

<sup>4</sup> For more information about ANOVA, see Anscombe (1948).

hoc tests confirmed existence of only three different groups which can be seen in Table 3, and homogenous categories are clustered together under each column. D2F stands for days remaining to flight and ChgClass is the price change class (integer between -3 to 3).

**Tests of Between-Subjects Effects**

Dependent Variable: Sales

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	3483.642 <sup>a</sup>	140	24.883	3.333	.000
Intercept	18985.967	1	18985.967	2543.115	.000
D2F	1338.382	20	66.919	8.964	.000
ChgClass	806.702	6	134.450	18.009	.000
D2F * ChgClass	1020.343	114	8.950	1.199	.077
Error	26846.420	3596	7.466		
Total	53554.000	3737			
Corrected Total	30330.062	3736			

Table 1: ANOVA results to test significance of price change and days to flight as main factors on sales

Significance of days remaining to flight departure and price change classes on sold seats is interpreted from Table 1, and no interaction between them exists.

In order to classify intensity of effect within each independent factor on sold seats, subsets with homogenous effect were identified, using a post-hoc Duncan's test<sup>5</sup>, which demonstrated ten homogenous subsets in days remaining to departure, and three homogenous subsets in price change classes. These subsets are increasingly ordered by their relevant average values of seat sales. Table 2 demonstrates homogenous classes in different days remaining to flight, for example in one, five and two days to flight departure the values of average sold seats are 3.76, 3.52, and 3.20 respectively and according to their variances, they don't demonstrate a significant difference and are classified in the 9<sup>th</sup> homogenous subset in Table 2.

---

<sup>5</sup> This post hoc test (or multiple comparison tests) is used to determine the significant differences between group means in an analysis of variance setting, (see Duncan (1955) for more information).

Duncan<sup>a,b,c</sup>

D2F	N	Subset									
		1	2	3	4	5	6	7	8	9	10
19	178	1.61									
18	178	1.64	1.64								
15	178	1.88	1.88	1.88							
20	177	1.89	1.89	1.89							
17	178	1.89	1.89	1.89							
12	178	2.06	2.06	2.06	2.06						
11	178	2.12	2.12	2.12	2.12	2.12					
16	178	2.13	2.13	2.13	2.13	2.13					
13	178	2.16	2.16	2.16	2.16	2.16					
14	178		2.30	2.30	2.30	2.30					
9	178			2.35	2.35	2.35					
4	178			2.35	2.35	2.35					
10	178			2.52	2.52	2.52	2.52				
8	178				2.57	2.57	2.57	2.57			
3	178				2.60	2.60	2.60	2.60			
7	178					2.77	2.77	2.77			
6	178						3.12	3.12	3.12		
2	178							3.20	3.20	3.20	
5	178								3.52	3.52	3.52
1	178									3.76	3.76
0	178										3.90
Sig.		.108	.052	.066	.120	.059	.064	.051	.196	.069	.223

Means for groups in homogeneous subsets are displayed.  
Based on observed means.  
The error term is Mean Square(Error) = 7.466.

a. Uses Harmonic Mean Sample Size = 177.952.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = .05.

Table 2: Homogenous subsets in days remaining to flight departure in terms of average values of sold seats in each subset

Similarly, the first subset in Table 3, with minimum average of sold seats, which is 1.75 belongs to the no-price-change class; the average sales subset includes: price decreases of more than 15%, and price increases of over 30%; high sales happens for price decrease less than 15% and more than 30% as well as price increases up to 30%.

Duncan<sup>a,b,c</sup>

ChgClass (% of price)	N	Subset		
		1	2	3
0	1269	1.75		
[-15%,-5%)	370		2.53	
[-100%,-15%)	616		2.81	2.81
(15%,100%]	602		2.86	2.86
(5%,15%]	364			3.03
[-5%,0)	265			3.04
(0%,5%]	251			3.16
Sig.		1.000	.104	.106

Means for groups in homogeneous subsets are displayed.  
Based on observed means.  
The error term is Mean Square(Error) = 7.466.

- a. Uses Harmonic Mean Sample Size = 405.089.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 3: Homogenous subsets in price change classes in terms of average sales

The results demonstrate the effectiveness of price change regardless of its direction on sales, the significance of price change effect in Table 1, and the same class of seat sales for negative and positive price changes in Table 3 are evidence of this statement. Also Table 1 might show some interaction between price change and days remaining to departure, although it is not significant; days remaining to flight also demonstrate significant effect on sales.

### 4.3. Conclusion

A significant effect of time remaining to flight on the amount of sold seats was found, which is not a new finding (see, e.g. Koenigsberg et al. (2008) and Sa (1987)). Although the effect of price change was identified as an important factor for sales forecasting; these findings confirm that the price change is an important factor on customers' decisions to buy a ticket.

Moreover, the lack of a significant interaction between price change and time, demonstrated the importance of price change, regardless of time. One of the approaches in Perishable Asset Revenue Management (PARM) studies in the airline industry is the difference in arrival time of high and low valuation customers; the last two statements imply that price change affects high valuation and low valuation customers equally.

## 5. Roundtrip models

Addition of a new variable into a model should follow a logical procedure. Basically there should be some relevance between target variable and independent variable. Such a relation can be revealed through studying the processes that affect the target variable. Furthermore, a statistical relationship between the independent and target variable should exist. In this section, we uncover a hidden relation between different flights, based on the idea of round trips.

Short staying periods at destination (less than three nights in average for local travels, see Webber et al. (2004), Table 2.2) is the motivation for proposing a new model to include factors such as *price* in both directions of a round trip. To the extent of the author's knowledge, return flights have not yet been used to improve sales forecasts.

In the existing models, *time remaining to a flight departure* and *price* are considered important factors, but in the proposed model, the *round-trip price* of tickets should be taken into account, instead of a *single direction's price*. If the passenger level information exists, the important returning flight(s) can be identified and the major round-trip prices are in hand, but in the absence of this data, the primary challenge becomes finding the returning flights for a certain flight, which will be called as linked flights with a certain flight.

The challenge of forecasting sales, using round-trip variables, can be translated into finding a relation between traffic in one direction and traffic in reverse direction, to which we refer as linked days, and

further identifying linked flights on these days. Finding linked days in a route can be counted as a confirmation for the existence of a balance in the number of travelers in two directions of the routes in this study. In subsection 5.1 and 5.2, the linked days and linked flights is studied.

One may think that travelers choose available return flights with an equal chance in/around their return day, although this study demonstrates more specific return flights can be considered for each flight in the other direction. This helps to achieve a better accuracy for seat sales in the round-trip forecasting models, comparing with one way sales models.

### **5.1. Linked days**

If the traffic volume of a certain day in one direction between a pair of cities could be explained by the reverse traffic volume of one or more days, we refer to such days as linked days. Finding this relation helps to understand the percentage of the travelers that are planning to stay for certain number of days at each end of a route. This subsection starts with demonstration of the traffic volume in two directions of the three routes to provide a basic visual assessment; it is continued with finding linked days and proposing a model for it. Using the model's coefficients, the percentage of passengers with different tendencies, in terms of trip duration is shown. In this subsection, the benefit of the mixture models is also explained which are employed for identifying the relation between the traffic in two directions of a route; mixture models will also be used in the next subsection to identify the linked flights.

Figure 4 demonstrates the traffic in the three routes in two directions for a randomly selected one month period. A bar for each day indicates traffic in one direction and the line graph demonstrate the traffic volume in reverse direction. Values on the vertical axis represent the total seats sold for all flights on a selected day.

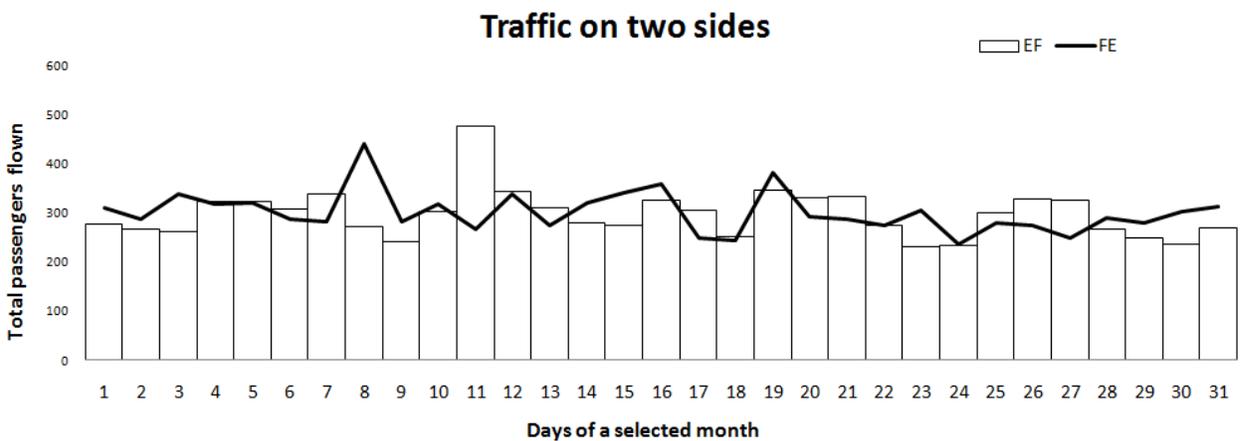
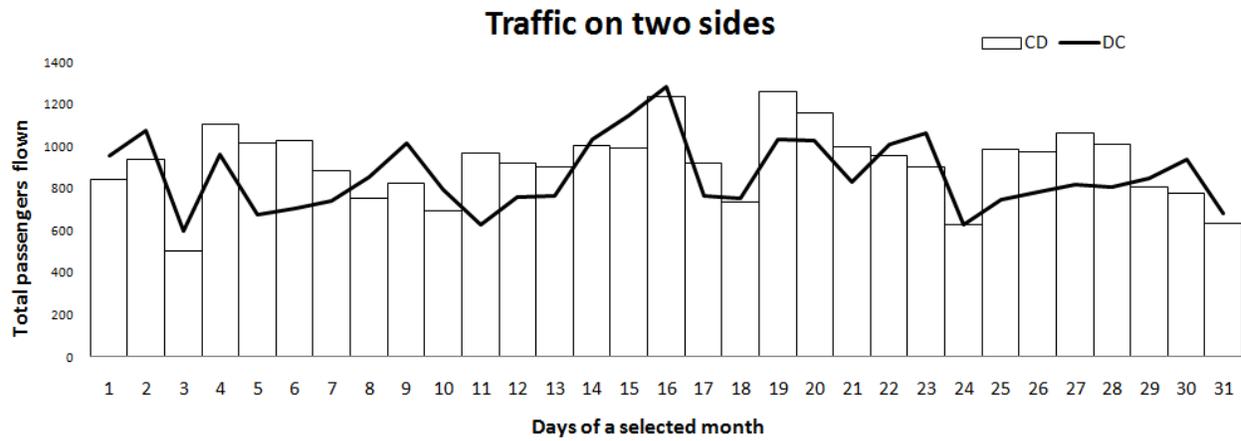
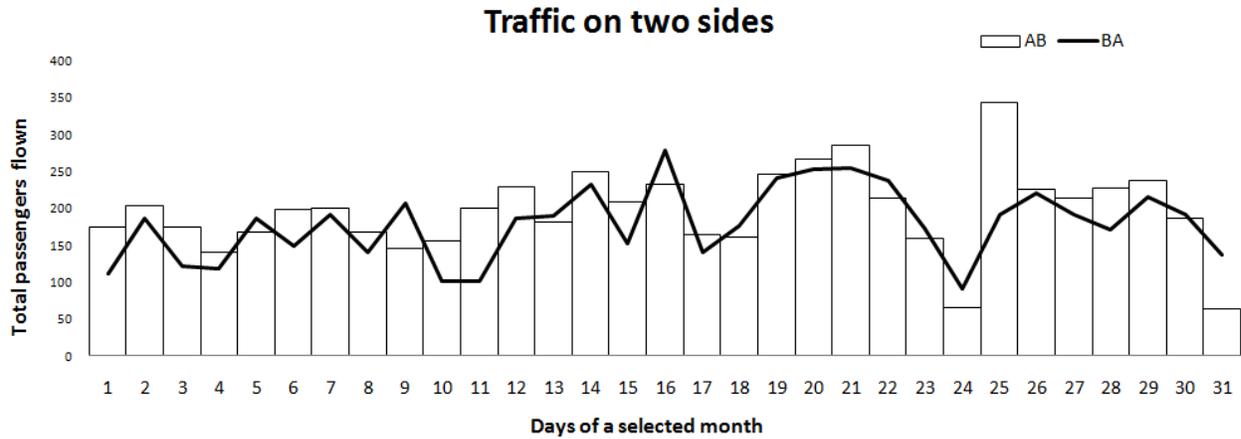


Figure 4: Enplanements in two directions in a randomly selected month for the three routes

Visual assessment of figure 4 reveals no evident relation between enplanement in the two directions. The only clear comparison between the traffic volumes is that they are closed together and it can be concluded that whether the people that travel in one direction, use the same airline to return, or at least the total seats sold in the two directions are very close together in any one week period.

Sold seats' data is not available at passenger level, so the relation between traffic in the two directions on each route should be found in another way. Our solution is constructing a model with daily enplanement in one direction as dependent variable, and the reverse enplanement in the neighbouring days as independent variables.

The possibility of having different segments of customers in terms of their trip duration is an important issue to be considered. Assume, e.g., there are two segments in travelers, a business group who return at the same day that are 30% of the total travelers, and the leisure travelers (the remaining 70%) that return after one week. If the leisure travelers would be reduced for example because of a bad weather and no separate model exists to estimate the enplanement of each group, then it can't provide a correct estimate for the enplanement in the reverse direction for today or enplanement one week from now. As a result, identifying segments within travelers lead to a better insight about them and provides the opportunity to study each segment separately. The Mixture models can be employed to identify the percentage of the data in each segment and finding models to describe the relation between target variable and independent variables in separate segments.

To identify an optimal number of segments, Bayesian Information Criterion (BIC) is used as an approximation to Bayesian factors (see Millar, 2006). BIC is basically a selection criterion of a model among a number of models of different dimensions. BIC resolves the problem of overfitting, which happens as a result of adding more parameters into a model by identifying the most suitable dimension (number of variables) in a model. By adding more parameters, better precision might be achieved, but the model loses the generality. This might be the case when a mixture model is proposed; the rule to

select the best model or number of segments is selecting the model with the minimum BIC value among the proposed models<sup>6</sup>.

Although it is not likely to find more than two or three segments within customers, the possibility for having between one and four segments is examined. The BIC for each of these segments is computed three times because the Expectation Maximization (EM<sup>7</sup>) algorithm converges only to the next local maximum of the likelihood.

The dataset for the planned mixture model had enplanements for a certain day, and 17 days of enplanement in the other direction (the 17 days were selected from eight days before, eight days after, and the same day enplanements in the other direction). The reason for selecting 17 days with 8 days before and after the selected date is because the average duration of stay for travelers inside Canada is around 3 nights (see Webber et al. (2004), Table 2.2). Maybe it can be considered that they stay 3 nights and 4 days, so selecting 8 days on each direction covers the majority of travelers. Also the reason for selecting 8 days on each side is because of the probable difference in staying period at the each end of a route.

To identify number of segments, BIC value is computed for all the three routes and the results are illustrated in Figure 5, which indicates the existence of only one segment in customers for all the routes.

---

<sup>6</sup> For more information about BIC, see Schwarz (1978).

<sup>7</sup> For more information about EM algorithm and its applications, see Dempster, Laird and Rubin (1977).

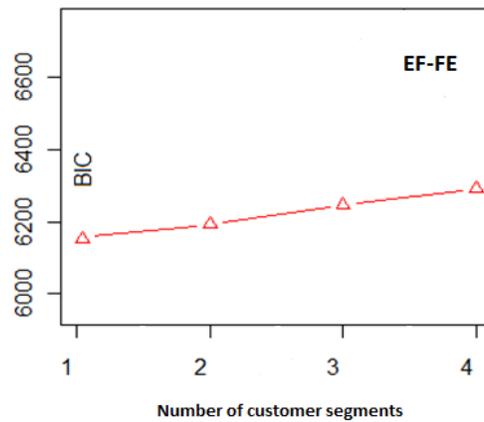
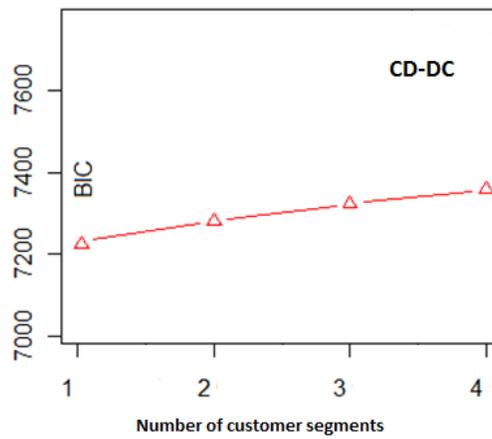
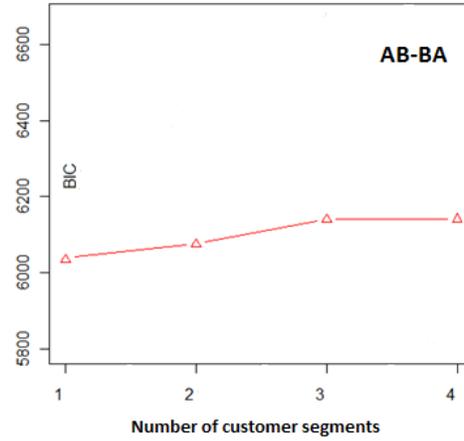


Figure 5: BIC graphs for selecting number of segments in each route

Having only one segment in customers, demonstrates that a mixture model is not needed in this case and a simple regression model may explain the traffic relation between the two directions.

As it was discussed in the beginning of this section, the coefficients of the independent variables in the regression model provide possibility to compare the share of travelers from neighbouring days or possibility to compare travelers with different duration of staying in each of the ends of a route.

In the AB-BA route, the significant coefficients at 95% confidence level are identified to be the coefficients for day seven and day eight before, as well as day five and day seven after, and also the same day. The interesting point is the significance of day eight before, which has a negative value, and means that the more request on that day results in less demand in the other direction for the target day. One explanation is considering a constant number of aggregated volumes of travelers in day seven and eight together and assuming consistency for the duration of the travel between the two cities. In other words, the travelers seem to be flexible for choosing one day before, although their travel duration remains constant (1 week).

In CD-DC route, the days with a significant effect on total sales were days 4, 5, 7, and 8 before (with similar effect of day 7 and 8 in AB-BA route) and days 1, 3, 4, and 7 after.

On EF – FE route, the only significant effect in days after is day three, the pattern of this route indicates that average trip length for travels from F to E is 4 days and it is around three days in the reverse direction.

The model summary, ANOVA and coefficient tables for all of the above linear models can be found in appendix-II.

The main finding of the above models is the identification of the average days of travel and special days that have a significant effect on the enplanement in the reverse direction. This helps to identify days with connected flights, although for developing a disaggregate model to predict sales, these relations should be studied at flight level.

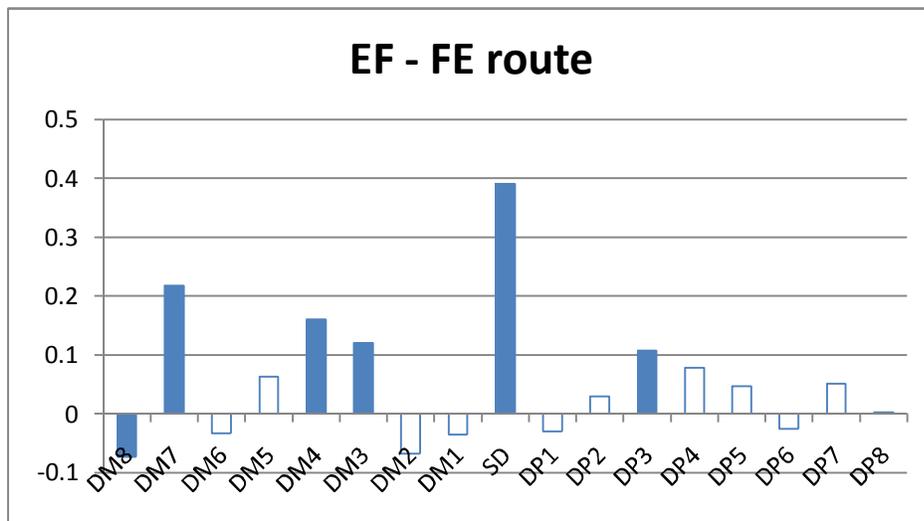
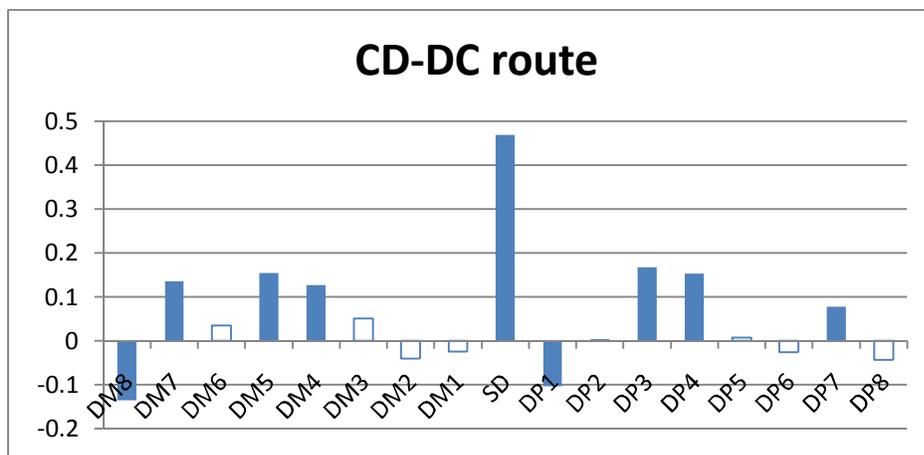
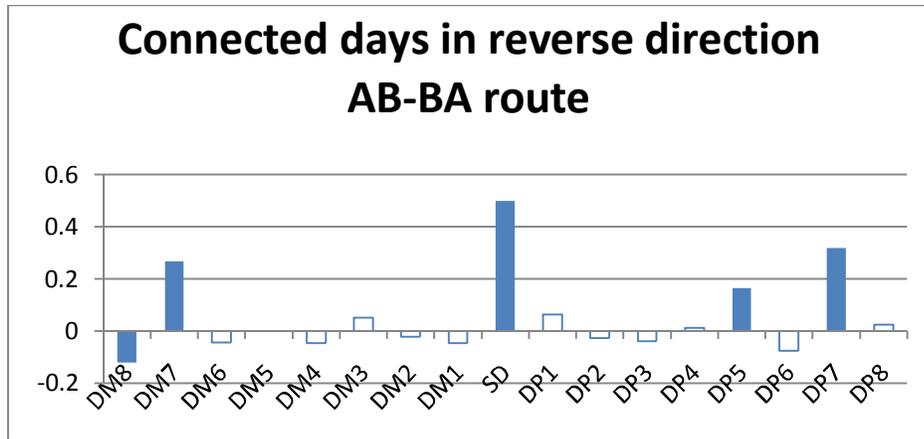


Figure 6: The traffic in 17 days on the reverse direction as independent variables are used to estimate a target day traffic; the coefficients of each in the resulting regression model are drawn; days with significant effect have filled bars.

## 5.2. Linked flights

All the flights that their demand significantly affects the demand of a flight in reverse direction are called linked flights. Linked flights help to construct models that consider aggregate costs of traveling in a round trip.

This section begins with a short explanation of the reason for selecting a limited set of flights and identifying linked flights only in AB-BA route; it continues with almost the same procedure in subsection 5.1 to identify linked flights, although with two segments in customers, a mixture model is developed which explains the percentage of people in each segment and their related preferences that is explained by their linear regression model.

The model in 5.1 relates traffic in one direction to the traffic in the reverse direction by employing 17 variables that each represents the aggregated sales on the same or on an adjacent day ( $[-8, +8]$ ). This number of variables was sufficient to cover all the traffic in the eight adjacent days before, after, and in the same day in reverse direction.

Because of limitation in our dataset and some missing data on each route, a dataset with no missing data that covers all the flights in a model became very small, as a result, the model variables should be reduced in this study. In order to find segments in the customers for choosing a return flight in the past or next 'n' days, a model is required whose number of variables is the result of multiplying 'n' by the number of flights in one day (i.e. for 'k' flights per day, 'nk' flights or variables should be in the model); hence a route with a minimal number of flights allows to maximize the span of days in the model. By choosing the AB-BA route and a range of two days, the model includes 13 flights in the other direction.

In the absence of passenger level data, linked flights need to be identified with mathematical modeling. Usually in a selection process, when human decision making is involved, different segments might exist, so in identification of customers' preferences for selecting a return flight, it is plausible to consider this point. To investigate existence of such segments within customers, mixture models are proposed and

BIC is used to identify the best number of segments. In subsection 5.1 the same method was used to identify linked days and in this subsection, it helps to identify linked flights.

The BIC values for one, two, and three segments are computed; with starting from one segment, one segment is added each time and BIC value is computed. The result is illustrated in Figure 7, which indicates a minimum value for BIC when there are two segments in customers.

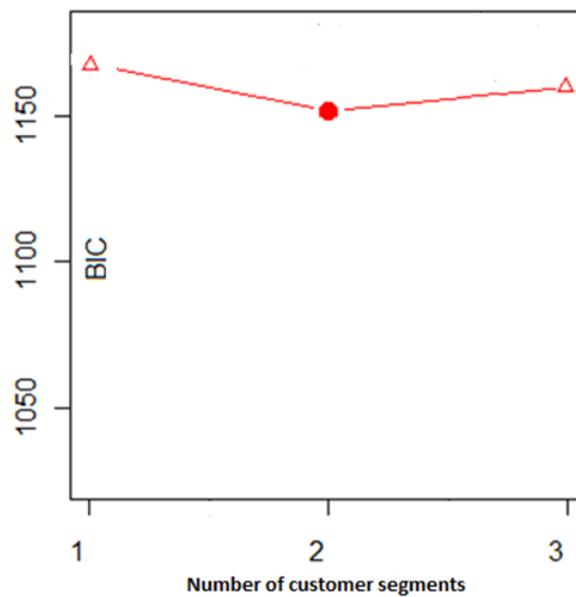


Figure 7: BIC values for one, two and three segments in the returning enplanements

Visual evaluation of the segment structure can be done using histograms or rootograms of the posterior segment probabilities Tantrum, Murua, and Stuetzle (2003). The rootograms<sup>8</sup> in Figure 8 demonstrate a clear separation between the two segments, because of a high concentration of posterior probabilities around the one and zero probabilities.

---

<sup>8</sup> Rootograms are very similar to histograms, the only difference is that the height of the bars correspond to square roots of counts rather than the counts themselves, hence low counts are more visible and peaks less emphasized. A peak at probability 1 indicates that a mixture segment is well separated from the other segments, while no peak at 1 and/or significant mass in the middle of the unit interval indicates overlap with other components Leisch (2004).

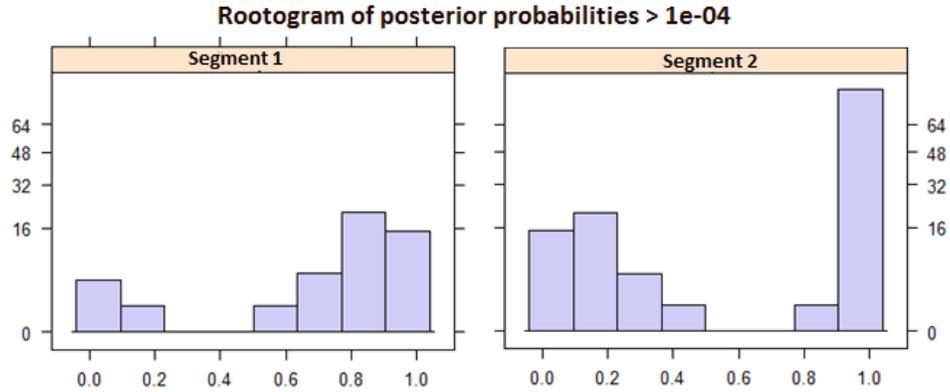


Figure 8: Rootograms for the two segments identified in return flight choice

To compute BIC values, the Expectation Maximization (EM<sup>9</sup>) algorithm is used; as this algorithm converges only to the next local maximum of the likelihood, it should be run repeatedly using different starting values, the algorithm for this problem is run 15 times. Using the mixture modeling, the linear models' coefficients of the two segments' are identified in Table 4, and the data is divided into two segments with 28% and 72% respectively as demonstrated in Figure 9. ('Comp' or component in Table4 is a notation in Flexmix library<sup>10</sup> for a segment).

```

                prior size post>0 ratio
Comp.1 0.28      42      47 0.894
Comp.2 0.72      88     130 0.677

'log Lik.' -500.3066 (df=31)
AIC: 1062.613   BIC: 1151.507

```

**Table 4: identification of the two segments and percentage of data in each segment**

The coefficients' analysis report is provided in Table 5, for the two segments. The significant coefficients in the larger segment are enclosed in a box in Table 5. They are also identified by stars which shows their significance level, as it is explained beneath the table 5.

<sup>9</sup> EM is an iterative optimization method to estimate an unknown parameter, given a measurement data, however there is a "hidden" nuisance variable, which needs to be integrated out; See the technical report by Dellaert (2002) and also Dampster, Laird and Rubin (1977).

<sup>10</sup> The 'R' is an open-source software for statistical analysis; R and its libraries (such as Flexmix) are free for research and the source code of its models and analysis tools are available. Of course the 'R' is not intended to be the primary reference for understanding the logic of the analysis in this research and additional sources are introduced respectively.

\$Comp.1					\$Comp.2				
	Estimate	Std. Error	z value	Pr(> z )		Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-53.632483	2.262314	-23.7069	< 2.2e-16 ***	(Intercept)	25.5545747	19.7397763	1.2946	0.195468
AfterN	-0.107645	0.022290	-4.8292	1.371e-06 ***	AfterN	0.3746844	0.1357587	2.7599	0.005781 **
Af1	-0.317151	0.017810	-17.8073	< 2.2e-16 ***	Af1	0.2787597	0.1239093	2.2497	0.024468 *
MD1	0.244442	0.025331	9.6500	< 2.2e-16 ***	MD1	0.0606114	0.1317925	0.4599	0.645588
Mo1	-0.210278	0.018642	-11.2799	< 2.2e-16 ***	Mo1	0.0141743	0.1021342	0.1388	0.889623
Af2	0.146192	0.019800	7.3835	1.542e-13 ***	Af2	0.2044481	0.1374235	1.4877	0.136824
MD2	-0.430206	0.019828	-21.6974	< 2.2e-16 ***	MD2	-0.0056436	0.1398683	-0.0403	0.967815
Mo2	0.023456	0.016884	1.3893	0.1648	Mo2	-0.0279152	0.1055847	-0.2644	0.791482
A1	0.902218	0.016077	56.1194	< 2.2e-16 ***	A1	0.0613639	0.1212860	0.5059	0.612896
M1	0.909015	0.032292	28.1497	< 2.2e-16 ***	M1	-0.1695533	0.1473714	-1.1505	0.249931
O1	-0.230823	0.018784	-12.2885	< 2.2e-16 ***	O1	0.2462173	0.1200416	2.0511	0.040257 *
A2	0.743576	0.015531	47.8778	< 2.2e-16 ***	A2	-0.2146462	0.1264923	-1.6969	0.089713 .
M2	0.541671	0.024214	22.3700	< 2.2e-16 ***	M2	0.1431445	0.1267227	1.1296	0.258650
O2	-0.354136	0.018012	-19.6614	< 2.2e-16 ***	O2	-0.0187651	0.1129354	-0.1662	0.868033
---									
			Signif. codes:	0 '***' 0.001	'**' 0.01	'*' 0.05	'.' 0.1	' ' 1	

Table 5: Analysis for significance of coefficient of each variable in two customer segments

In the first segment which has 28% of data, most of the flights demonstrate a significant effect, while the second segment with the majority of the data(72%), demonstrates significance of only three flights, including: afternoon flight on the same day, the afternoon flight in one day before, and the next morning flight. As a result of this analysis, afternoon flight in the AB-BA route is one of the linked flights with the morning flight on the other direction. This flight is used to build the round-trip model in the next section; the reason for selecting only one flight in this modeling is data limitation.

Identifying few flights as specific returning flights demonstrated that it is plausible to think about Round Trip Models (RTM) for forecasting seats sold. It showed that in contrast with this thought that people may return with any flight, they tend to return with specific flights in a route and if this could be identified in a route, like our example, there is the opportunity for a better revenue management, using linked flights. This possibility depends on investigating the performance of RTM versus one way forecasting models which will be explored in the next section.

## **6. Fuzzy forecasting**

Fuzzy systems are about evaluating the relative membership of an input or output in a series of sets and establishing a rules-based system which maps an input dataset into an output dataset; these rules can be defined when constructing a decision making model in base of desired rules or they can be identified, when trying to investigate the relation between an input and output dataset. Adaptive Neuro-Fuzzy Inference Systems (ANFIS) are the kind of fuzzy systems that identify the rules and relations between input and output datasets and will be used in this section to develop the required forecasting models. Fuzzy systems can be used in many situations, such as decision making, systems' control or forecasting, (see Becher (2008), Kuo, Wu and Wang (2002), and Kuo and Xue (1998)).

This section starts with explaining general concept of fuzzy logic and fuzzy systems' applications, as well as providing references for more information about the computations and models which are used in this section; References to fuzzy systems with application in revenue management, airline industry, decision making, and forecasting problems are also provided. It proceeds with development and comparing price and price change models, as well as comparing a one way and a round trip model for forecasting seats sales.

### **6.1. Advantages of using a fuzzy model**

Fuzzy models are suitable when there is uncertainty in the data, imprecision in the input or output datasets, and complex interactions among variables. In such cases, fuzzy systems can be more suitable and can manage the complexities in the models, comparing with models that use a crisp logic<sup>11</sup>.

The “father of fuzzy”, Zadeh, believes that the fuzzy logic is a precise mathematics to explain imprecision. Also he says: “As complexity rises, precise statements lose meaning and meaningful statements lose precision.” (MathWorks 1995)

---

<sup>11</sup> The True and False logic is sometimes called the crisp logic.

The grey logic<sup>12</sup>, tries to relate a variable's value into different sets, instead of previous logic of assigning a value only to one set; for example, a flight ticket can be classified with some degree of expensiveness and with some degree of being cheap, rather than either being identified as expensive or cheap.

The challenge in this kind of modeling is identifying the interaction between variables, defining the necessary classes within variables, establishing rules, and assigning suitable weight for each rule to map the input dataset into the output datasets. With recent advancement in the field, some of these steps can be done automatically using adaptive methods. In this approach all the possible rules between input and output variables' classes are generated and a neural network identifies the weight of each rule and the shape of each class within a variable<sup>13</sup>. This makes fuzzy systems much more convenient for modeling, comparing with modeling methodologies which require testing different arrangements of independent variables in a model to estimate a target variable.

For more on the field of fuzzy logic, see (Zadeh 1973), also detailed explanation about fuzzy systems and their application in revenue management are provided by Becher (2008), and the adaptive methods in fuzzy systems are covered in MathWorks (1995).

Previous research in different disciplines has confirmed the preference of fuzzy models over common models such as regression, which is reflected in research of Kuo and Xue (1998) or Kuo, Wu and Wang (2002) that have applied fuzzy method specifically for sales forecasting. There are other references that were discussed in literature review and we will not repeat them here again.

According to the ability of the adaptive fuzzy systems in identifying complex relations between variables, and regarding researchers' preference in the field, it is used as our modeling approach.

---

<sup>12</sup> Fuzzy logic, as a logic beyond crisp logic of true and false, or black and white, sometimes is called grey logic.

<sup>13</sup> Only the number of classes within each variable still should be identified by the user. I have proposed a procedure to identify the minimum possible classes within each variable in appendix III, although I believe it might be still simplified using BIC to find the optimum number of classes for each variable.

## 6.2. Comparing the effect of price and price change in sales forecasting

In the study of price-change effects on sales (section 4), effective role of price-change was identified. The inspiration of effectiveness of price change came from observing higher sales volumes under price change with constant price situations. Of course *price change* is a derivative of *price* and *price* sometimes has some information that can not be derived from *price change*. However, if *price change* demonstrates a better performance to estimate the sales, then it might be a good replacement for price in a sales forecasting model, although price should exist as an additional constraint; therefore, two fuzzy models were developed for a comparative study. The first model uses the days remaining to flight, weekday and price as its factors and the other model uses price-change instead of price. Several different numbers of classes within input variables were tested and the best number of Member functions (MFs)<sup>14</sup> for the price model is derived from one MF for the days remaining to flight, two MFs for price itself, and three MFs for weekdays. In the price change model, the best number of MFs 2, 1, and 3 respectively. The comparative results of training, checking and testing for price and price-change model are demonstrated in Table 6.

	Price model	Price-Change(PC) model
Training error	2.0569	2.0398
Checking error	1.9969	2.0127
Error difference between Training and Checking	0.06	0.027
Testing error	2.1471	2.2383

Table 6: Comparing errors (RMSE<sup>15</sup>) for training and checking datasets in price and price-change model

To control for overfitting, a randomly selected checking-dataset is used. One may claim to have found a better model with a smaller RMSE using a training dataset, but if the distance of the training error and the checking dataset grows, it means that the model is not generalizable. As a result, a model with a smaller difference between the checking and training errors is expected to provide better forecasting

---

<sup>14</sup> To learn about selection criteria for the number of member functions, see appendix III.

<sup>15</sup> RMSE stands for Root Mean Squared Error, which is the error measure used in this thesis; please see the appendix I – error measures, for more information about its definition and the reason for its selection.

results. The distance between the training dataset and the checking dataset may happen because of too much iteration, which adapts the model parameters using training data; this also might be because of the independent variables' nature and their relation with the target variable.

The results demonstrate that the RMSE between the checking and training data in price-change model is half of the difference in the price model. This means that using the price-change model will more likely result in more accurate estimates; the randomly selected testing-dataset shows 4% less RMSE for the price model although it should not be judged the reason for superiority of price model over price-change model; however, it demonstrates that the results of both models are very close to each other.

Another idea in forecasting is to study the price change sequence. To do so, two consequent price-changes, which require three prices, are used to build another model. With two additional prices for the price model and one more price change for the price-change model, the following results were derived:

	Price model	Price-Change(PC) model
Training error	2.1514	2.1924
Checking error	2.4015	2.3985
Error difference between Training and Checking	-0.25	-0.201
Testing error	2.3112	2.354

Table 7: Comparing errors (RMSE) for training and checking datasets in price and price-change model for two consequent price change and three prices

The above results come from a different larger dataset than the dataset used in the previous model. The optimum number of member functions for the price model is derived from one MF for the days remaining to flight, one MF for each of the three prices, and two MFs for weekdays. These numbers for the price change model are one MF for days remaining to flight, one MF for each of the two price-changes and two MFs for weekdays of the transaction date.

In conclusion, price change models have slightly better results in terms of stability in predictions (less variance); they can also be used to reduce the number of variables when a group of consequent prices

are used in a model. The reasonable precision of price change models demonstrates that price change is as important as price in affecting the sales, a tool which might be used already by airlines; price volatility specially, in the last two weeks before flight departure might be an evidence of using this phenomenon.

### **6.3. Comparing the efficiency of one-way and round-trip models**

So far, a relation between the passengers' traffic in the two directions was identified, and the total seat sales for a flight were estimated by seats sold in linked flights. This subsection intends to introduce a round-trip model and compare them with one way models. Round-trip models for sales forecasting use the aggregate prices for a round-trip instead of having a single direction price data. Our hypothesis is that if there is a connection between enplanement of two or more flights, then such connection might exist at disaggregate level; to investigate the validity of this hypothesis, two models are studied:

- A one way model
- A model, with linked flights (the round trip model)

The hypothesis is supported if the round trip model demonstrates a relatively smaller prediction error. In order to select the independent variables for this model, price, price change and days remaining to flight were considered, as well as seasonal factors such as weekdays and months effects. Regarding the limitation in the available data, the models should be simplified enough to be able to train the fuzzy system with the available datasets. Some comparative experiments demonstrated less significance of month and price factors; to incorporate the price effect, price-change (which demonstrated a significant effect earlier in section 4) is used; weekdays also demonstrate a significant effect on seat sales and in several studies such as Sa (1987), Koenigsberg et al (2008) or Lee (1990) is used. Hence, the final selected independent variables for both models will be:

- *Price change*
- *Days remaining to departure*

- *Weekday* (of transaction)

In the second model, the *price change* for the selected linked flight is added as an additional independent variable. In the simple model that we built, only one linked flight is selected which is the flight in the afternoon of the same day and is paired with the morning flight. This flight demonstrated a large effect on estimating the sales in the other direction using the regression mixture model, as it is shown in Table 5.

The combination of variables and optimum number of member functions to build the two models are shown in Table 8.

One-Way Model Variables	Round-Trip Model Variables	Number of member functions
Days remaining to flight	Days remaining to flight	2
Price change in the morning flight	Price change in the morning flight	1
----	Price change in the returning afternoon flight	1
Weekday	Weekday	3

Table 8: List of variables & the relative number of member functions in one-way and round-trip models

In Table 9 the errors for the one-way and the round-trip models are summarized. The errors are Root Mean Squared Error of estimations. We selected RMSE as the measure for errors in this study, because we found some preferences for RMSE, when comparing with MAPE; see the appendix I for more explanations.

	One-way model	Round trip model	% One-way vs. Round-trip (O-R)/O
Training error	2.39	2.35	1.7 %
Checking error	2.76	2.63	4.7%
Testing error	2.30	2.23	3.0%
Error difference between Training and Checking	0.37	0.28	24%

Table 9: Comparative table for the one-way and the round-trip models' errors

## 6.4. Results

Two important points are associated with the results in Table 9. The first noticeable point is a 24% smaller difference between the training and the checking errors, which is an indication of a more reliable model; in other words, entering the additional variable has improved the model and has reduced randomness in estimation.

Moreover, the errors for training, checking and testing data reduced by 1.3%, 4.7%, and 3.0% respectively. It should be remembered that, in this concept model, the information of only one linked flight is used and it is expected that we would observe smaller error values by having more linked flights in the model.

Another way to improve the model's reliability and performance is by more accurately identifying of the linked flights. What was done in subsection 5.2 was guessing the linked flights, which makes the model more prone to errors. This weakness can be eradicated by using detail information at the passenger level.

With a look at Table 6, there are at least two more linked flights in the second segment and around 13 linked flights in the first segment that could be used in a model to decrease the error. The testing error has reduced 3% by introducing only one variable, and it is reasonable to consider more improvement by incorporating additional linked flights information. Research by Weatherford and Belobaba (2002) demonstrated a 1-4% increase in revenue for each 10% reduction in the forecasting error. It is expected that this research will help the passenger airlines businesses to better forecast and to manage their resources more efficiently.

## 7. Conclusion and further research

This thesis covered some possibilities to improve sales forecasting models in passenger airlines. The analysis of enplanement in two directions of a route demonstrated the existence of a connection between them; this result led to find a connection between the enplanement at the flight level (for flights in reverse directions). This finding again evoked possibility of using aggregated price of the flights with connected enplanement to improve sales forecasting model; this idea was supported by finding a higher precision for Round Trip Models (RTM) versus one-way forecasting models.

A potential benefit of studying round trip patterns is the vision it provides for a pricing policy on a route. There is a potential to maximize revenue in a route while being competitive with a few prices to capture a target market instead of being competitive in all flights' prices on that route. Although this subject is not covered in this research, it remains as an opportunity for further research in RTMs.

Another finding in this thesis is the significant effect of price change and insignificance of its interaction with the days remaining before a flight, which provides an opportunity to plan for price variations in order to gain higher revenue, not only in the last remaining days before a flight, but also during other sales period. The precision of price models was a little bit lower than price-change models and price change models demonstrated less variance in their estimates, which guarantee more reliability in predicted values.

Adaptive Neuro-Fuzzy Inference Systems, (ANFIS) were used in developing forecasting models and its benefits were mentioned, however if BIC could be applied for finding optimum number of membership functions, adaptive fuzzy models might be improved. Also fuzzy mixture models could be used to increase the forecasting precision. For example in our study, fuzzy mixture model might improve forecasting by providing separate models for the two segments of travelers for returning flight choice.

ANFIS or other adaptive models are the kind of models that are often criticized because they need a lot of data for training, they are not suitable for ad-hoc decision making, and they are not responsive to

rapid changes. The author suggests developing models with different dataset sizes to try to study the variation of error range as input dataset size varies. Identifying the minimum number of records which are required for an acceptable level of precision provides a balance between precision and reliability in prediction results.

Another idea for future work is applying fractal models to forecasting, which has not yet been used in revenue management. These models have demonstrated better results in financial applications, when high degree of non-linear variability exists (see Richards (2010)), hence it may work for forecasting sales in revenue management and specifically in the passenger airline industry.

The relation between change of precision in forecasting and revenue change is studied by Lee (1990) and by Weatherford and Belobaba (2002). The author believes that in their research, they had similar sets of data in terms of percentage of sold seats, so it is suggested to repeat this study for situations with smaller sold seat percentages. In other words, a future study could explain revenue changes regarding improvement in forecasting precision in terms of percentage of sold seats, i.e. between two flights with 60% and 99% of sold seats, there might be some difference in the effect of higher precision in forecasting on revenue, because higher precision in the 60% sold seats situation may improve sales, however when 99% of seat sold, this will help to increase revenue with selling at higher prices for the same amount of sold seats, which is a percentage of the seat price.

Another important issue which needs attention of researchers in the field is forming a group to collect and share anonymous data. It is also suggested to collect specifically the price information from travel websites which creates better research possibilities. i.e. if in this research, the rivals prices were available, their effect on sales could be studied and forecasting models could be improved. It can help for future studies, if a group of researchers could make a small investment on any web-crawler software, and the required storage on a web server for collecting and sharing the price data.

## References

- Anscombe, F. J. "The validity of comparative experiments." *Journal of the royal statistical society, series A (General)* (Blackwell publishing for the royal statistical society) 111, no. 3 (1948): 181-211.
- Becher, Michael. *Integrated Capacity and Price Control in Revenue Management*. Wiesbaden: Gabler, (2008).
- Beckmann, M. J., and F. Bobkowski. "Airline Demand: An Analysis of Some Frequency Distributions." *Naval Research Logistic* 5 (1958): 43-51.
- Belobaba, Peter P. "Comparing decision rules that incorporate customer diversion in perishable assets revenue management situations." *Decision Sciences* 27, no. 2 (1996): 343-363.
- Bitran, Gabriel, and Rene Caldentey. "An overview of pricing models for revenue management." *Manufacturing and service operations management* (Manufacturing and service operations management) 5 (2002): 203-229.
- Dampster, A. P., N. M. Laird, and D. B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the royal Statistical society, series B (Methodological)* 39, no. 1 (1977): 1-38.
- Dellaert, Frank. *The expectation maximization algorithm*. Technical, College of computing, Atlanta, Georgia: Georgia institute of technology, (2002).
- DeSarbo, WS, and M Wedel. "A mixture likelihood approach for generalized linear models." *Journal of Classification* 12 (1995): 21-55.
- Duncan, B. D. "Multiple range and multiple F tests." *Biometrics* 11 (1955): 1-42.
- Farias, Vivek Francis. "Revenue management beyond "estimate then optimize"." *Ph.D dissertation*. Stanford, California: Stanford University, (2007).
- Feng, Youyi, and Baichun Xiao. "A continuous-time yield management model with multiple prices and reversible price change." (INFORMS) 46, no. 5 (2000): 644-657.
- Feng, Youyi, and Guillermo Gallego. "Perishable asset revenue management with Markovian time dependent demand intensities." *Management Sciences* (INFORMS) 46 (2000): 941-956.
- Gallego, G., and G. J. van Ryzin. "A multi-product dynamic pricing problem and its applications to network yield management." *Operations Research* (Operations Research) 45 (1997): 24-41.
- Gillen, David, and Benny Mantin. "Price Volatility in airline markets." *Transportation Research Part E* (Logistics and transportation review) 45, no. 5 (2009): 693-709.
- Kamakura, WA, and M Wedel. *Market Segmentation - Conceptual and Methodological*. 2nd . Boston, MA: Kluwer Academic Publishers, (2001).
- Koenigsberg, Oded, Eitan Muller, and Naufel J. Vilcassim. "easyJet® pricing strategy: Should low-fare airlines offer last-minute deals?" *Quantitative Marketing and Economics* 6, no. 3 (2008): 279-297.

- Kuo, R J, P. Wu, and C. P. Wang. "An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination." *Neural Networks* 15, no. 7 (2002): 909-925.
- Kuo, R. J., and K. C. Xue. "A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights." *Decision Support Systems* 24, no. 2 (1998): 105-126.
- Lee, Anthony Owen. "Airline Reservation forecasting: Probabilistic and Statistical Models of the Booking Process." *Ph.D. Thesis*. MA: MIT, (1990).
- . "Airline Reservation forecasting: Probabilistic and Statistical Models of the Booking Process." *Ph.D. Thesis*. MA: MIT, September (1990).
- Leisch, Friedrich. "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R." *Journal of Statistical Software* 11, no. 8 (2004): 1-18.
- Levina, Tatsiana, Yuri Levin, Jeff McGill, and Mikhail Nediak. "Dynamic pricing with online learning and strategic consumers:an application of the aggregating algorithm." *Operations research* 57, no. 2 (2009): 327-341.
- Littlewood, Ken. "Forecasting and control of passenger bookings." *Journal of Revenue and Pricing Management* 4 (2005): 111-123.
- Mantin, Benny, and Bonwoo Koo. "Dynamic price dispersion in airline markets." *Transportation Research Part E* 45 (2009): 1020-1029.
- MathWorks. *Fuzzy Logic Toolbox User's Guide*. Natick, MA, (1995).
- McGill, J. I., and G. J. Van Ryzin. "Revenue Management: Research overview and prospects." *Transportation Science* 33, no. 2 (1999): 233-253.
- McGill, Jeffrey. "Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints." *Annals of Operations Research* , (1995): 209-240.
- McLachlan, G, and D Peel. *Finite Mixture Models*. John Wiley and Sons Inc., (2000).
- Millar, Paul. "Using the Bayesian Information Criterion to judge models and statistical significance." *North American stata users' group meeting*. Boston, Alberta, (2006).
- Nam, K., and T. Schaefer. "Forecasting international airline passenger traffic using neural networks." *Logistics and Transportation Review*, (1995): 239-252.
- Profillidis, V. A. "Econometric and fuzzy models for forecast of demand in the airport of Rhodes." *Journal of air transport management* 6 (2000): 95-100.
- Richards, R. Gordon. "A fractal forecasting model for financial time series." *Journal of forecasting*, (2010): 586-601.
- Sa, Joao. "Reservations forecasting in airline yield management." *Master thesis*. MA: MIT, (1987).

- Schwarz, Gideon. "Estimating dimension of a model." *The annals of Statistics* 6, no. 2 (1978): 461-464.
- Talluri, Kalyan T., and Garrett van Ryzin. *The theory and practice of revenue management*. New York: Springer, (2005).
- Talluri, Kalyan, and Garrett van Ryzin. *The theory and practice of revenue management*. New York: Springer Science+Business Media, (2005).
- Taneja, Nawal K. *Airline Traffic Forecasting*. D.C. Heath and Company, (1978).
- Tantrum, J, A Murua, and W Stuetzle. "Assessment and Pruning of Hierarchical Model Based Clustering." *the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM Press, (2003): 197-205.
- Weatherford, L. R., and S Pöhl. "Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues." *Journal of Revenue and Pricing Management* , (2002): 234.
- Weatherford, L. R., T. W. Gentry, and B. Wilamowski. "Neural network forecasting for airlines: A comparative analysis." *Journal of Revenue and Pricing Management* 1, no. 4 (2003): 319-331.
- Weatherford, Lawrence R., and Samuel E. Bodily. "A taxonomy and research overview of perishable-asset revenue management: yield management, overbooking, and pricing." (INFORMS) 40, no. 5 (1992).
- Weatherford, R. L., and P. P. Belobaba. "Revenue impacts of fare input and demand forecast accuracy in airline yield management." *The journal of the operational research society* 53, no. 8 (2002): 811-821.
- Webber, Maryanne. *Canadian travel survey - domestic travel 2002*. Annual, Ottawa: Statistics Canada, (2004).
- Zadeh, L A. "Outline of a new approach to the analysis of complex systems and decision processes." *IEEE Transactions on Systems, Man, and Cybernetics* 3, no. 1 (1973): 28-44.

## Appendix I - Error Measures

Defining error measures is crucial to any comparative research. In order to have comparable results with previous research, similar error measures should be selected. In literature review, two measures of forecast accuracy were identified: mean absolute percentage error (MAPE), and root mean square error (RMSE), as used by Lee (1990), and Weatherford, Gentry, and Wilamowski (2003).

The mean absolute percentage error is defined by

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|,$$

where 'n' is the number of cases,  $A_t$  is the actual value, and  $F_t$  is the forecasted value. This measure has some shortcomings. One of them is that when  $A_t$  is zero, it is not defined which happens in most of the cases in demand forecasting. To avoid error we should omit these cases. In the models, which try to minimize error from average, such as regression, omitting the zero cases will increase the models' performances because a zero value is in one of the extents from the average when target value is positive (sold seats are non-negative). MAPE is the main error measure used in the comparative study of Weatherford, Gentry, and Wilamowski (2003) to compare regression models with neural networks in form of percentage error.

Another common error measure is RMSE which is defined below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2},$$

In sales forecasting research, this error measure is also used, see Lee (1990). In this study we use this measure because it doesn't ignore zero values like MAPE and provides a better comparison between models developed using different methods. As an example, in an effort for forecasting sales, using a regression and a fuzzy model, the two models had almost the same RMSE for the training dataset and regression model was showing a much smaller (better) MAPE. The reason was smaller variance in

estimations of the regression model; the results which were obtained in regression model were close to the training dataset average while the fuzzy model demonstrated some larger variance in estimated values.

## Appendix II - Linear models for connected days (subsection 5.1)

The model summary, ANOVA table and coefficients for AB- BA route

**Model Summary**

Model	R	R Square <sup>b</sup>	Adjusted R Square	Std. Error of the Estimate
1	.980 <sup>a</sup>	.961	.960	40.013

a. Predictors: DP8, DM8, DM4, DP4, SD, DM2, DP2, DP6, DM6, DM7, DP5, DP3, DM5, DM3, DP7, DM1, DP1

b. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

**ANOVA<sup>c,d</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.221E7	17	1306369.247	815.970	.000 <sup>a</sup>
	Residual	902964.798	564	1601.001		
	Total	2.311E7	581			

a. Predictors: DP8, DM8, DM4, DP4, SD, DM2, DP2, DP6, DM6, DM7, DP5, DP3, DM5, DM3, DP7, DM1, DP1

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Dependent Variable: Traffic from A to B

d. Linear Regression through the Origin

**Coefficients<sup>a,b</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	DM8	-.137	.055	-.120	-2.480	.013
	DM7	.307	.057	.268	5.343	.000
	DM6	-.051	.058	-.044	-.884	.377
	DM5	-.005	.056	-.004	-.085	.932
	DM4	-.053	.056	-.047	-.953	.341
	DM3	.059	.056	.052	1.066	.287
	DM2	-.025	.056	-.022	-.454	.650
	DM1	-.052	.061	-.046	-.867	.386
	SD	.570	.061	.500	9.351	.000
	DP1	.073	.060	.064	1.207	.228
	DP2	-.030	.056	-.026	-.535	.593
	DP3	-.044	.056	-.038	-.783	.434
	DP4	.014	.056	.012	.251	.802
	DP5	.187	.055	.165	3.392	.001
	DP6	-.085	.057	-.075	-1.481	.139
	DP7	.363	.057	.320	6.368	.000
	DP8	.027	.055	.024	.491	.624

a. Dependent Variable: Traffic from A to B

b. Linear Regression through the Origin

The model summary, ANOVA table and coefficients for CD-DC route

**Model Summary**

Model	R	R Square <sup>b</sup>	Adjusted R Square	Std. Error of the Estimate
1	.992 <sup>a</sup>	.983	.983	111.858

a. Predictors: DP8, DM3, DM7, DP5, DM5, DM1, DP3, DP7, DM8, DP1, DM4, DM2, DP2, DP4, DM6, DP6, SD

b. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

**ANOVA<sup>c,d</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.137E8	17	2.433E7	1944.769	.000 <sup>a</sup>
	Residual	7056852.415	564	12512.150		
	Total	4.207E8	581			

a. Predictors: DP8, DM3, DM7, DP5, DM5, DM1, DP3, DP7, DM8, DP1, DM4, DM2, DP2, DP4, DM6, DP6, SD

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Dependent Variable: Traffic from C to D

d. Linear Regression through the Origin

**Coefficients<sup>a,b</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	DM8	-.139	.034	-.135	-4.118	.000
	DM7	.140	.036	.136	3.879	.000
	DM6	.036	.036	.035	1.001	.317
	DM5	.159	.034	.154	4.638	.000
	DM4	.130	.034	.127	3.806	.000
	DM3	.052	.034	.051	1.521	.129
	DM2	-.041	.035	-.040	-1.189	.235
	DM1	-.026	.038	-.025	-.685	.493
	SD	.481	.039	.468	12.417	.000
	DP1	-.106	.038	-.103	-2.796	.005
	DP2	.003	.035	.003	.078	.938
	DP3	.171	.035	.167	4.950	.000
	DP4	.157	.034	.153	4.552	.000
	DP5	.008	.034	.007	.219	.827
	DP6	-.027	.036	-.026	-.748	.455
	DP7	.079	.036	.077	2.174	.030
	DP8	-.044	.034	-.043	-1.287	.199

a. Dependent Variable: Traffic from C to D

b. Linear Regression through the Origin

The model summary, ANOVA table and coefficients for EF-FE route

**Model Summary**

Model	R	R Square <sup>b</sup>	Adjusted R Square	Std. Error of the Estimate
1	.988 <sup>a</sup>	.976	.976	44.324

a. Predictors: DP8, DM7, DM1, DP2, DM4, DP5, DM3, DM6, DP7, DM8, DP3, DM2, DM5, DP4, DP6, DP1, SD

b. For regression through the origin (the no-intercept model), R Square measures the proportion of the variability in the dependent variable about the origin explained by regression. This CANNOT be compared to R Square for models which include an intercept.

**ANOVA<sup>c,d</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.589E7	17	2699122.645	1373.850	.000 <sup>a</sup>
	Residual	1108058.042	564	1964.642		
	Total	4.699E7	581			

a. Predictors: DP8, DM7, DM1, DP2, DM4, DP5, DM3, DM6, DP7, DM8, DP3, DM2, DM5, DP4, DP6, DP1, SD

b. This total sum of squares is not corrected for the constant because the constant is zero for regression through the origin.

c. Dependent Variable: Traffic from E to F

d. Linear Regression through the Origin

**Coefficients<sup>a,b</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	DM8	-.072	.036	-.073	-1.992	.047
	DM7	.215	.040	.218	5.419	.000
	DM6	-.033	.039	-.033	-.832	.406
	DM5	.062	.038	.063	1.612	.107
	DM4	.158	.039	.160	4.099	.000
	DM3	.118	.038	.120	3.061	.002
	DM2	-.067	.038	-.068	-1.748	.081
	DM1	-.034	.042	-.035	-.806	.421
	SD	.385	.043	.391	8.912	.000
	DP1	-.029	.042	-.030	-.688	.492
	DP2	.029	.039	.029	.751	.453
	DP3	.106	.039	.107	2.734	.006
	DP4	.076	.039	.078	1.968	.050
	DP5	.046	.039	.047	1.200	.231
	DP6	-.026	.040	-.026	-.648	.517
	DP7	.051	.040	.051	1.266	.206
	DP8	.003	.037	.003	.086	.932

a. Dependent Variable: Traffic from E to F

b. Linear Regression through the Origin

### **Appendix III – Selection criteria for the number of member functions in a fuzzy system**

The number of Member Functions (MFs) of a variable depends on its interaction with the other variables or its non-monotonic effect on target variable; in other words, the non monotonic effect of a variable on the target variable(s) regardless of other independent variables' values or under their effect must be explained with more than one member function; identifying the correct number of member functions is an important step in fuzzy models. To construct a stable and high precision fuzzy model, the difference between the training and the checking errors should be kept at the minimum value, and for this purpose, the type and the number of membership functions for each of the variables should be set properly. One approach for finding the proper number of the member functions is a gradual increase of the member functions for each variable while monitoring the changes in error differences between the training and the checking errors. If adding a member function reduces the error, another member function should be added, and if the difference between training and checking errors grows, the number of member functions should be reduced by one and stop the addition of a new member function to capture the interaction of the selected variable with other variables. Adding another member function also can be evaluated in terms of the processing time and a trade of between precision required and processing time may affect number of member functions.