Modeling of United States Airline Fares –

Using the Official Airline Guide (OAG) and

Airline Origin and Destination Survey (DB1B)

by

Krishna Rama-Murthy

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil Engineering

APPROVED BY:

Dr. Antonio A. Trani, Chair
Dr. Hojong Baik, Committee Member
Dr. Hesham Rakha, Committee Member

June 23, 2006
Blacksburg, Virginia

Keywords: Airline Fares, DB1B, OAG, Non-linear regression, Multiple linear regression, and Heteroscdacity.

Modeling of United States Airline Fares – Using the Official Airline Guide (OAG) and Airline Origin and Destination Survey (DB1B)

(Krishna Rama-Murthy)

ABSTRACT

Prediction of airline fares within the United States including Alaska & Hawaii is required for transportation mode choice modeling in impact analysis of new modes such as NASA's Small Airplane Transportation System (SATS). Developing an aggregate cost model i.e. a 'generic fare model' of the disaggregated airline fares is required to measure the cost of air travel. In this thesis, the ratio of average fare to distance i.e. fare per mile and average fare is used as a measure of this cost model. The thesis initially determines the Fare Class categories to be used for Coach and Business class for the analysis .The thesis then develops a series of 'generic fare models' using round trip distance traveled as an independent variable. The thesis also develops a set of models to estimate average fare for any origin and destination pair in the US. The factors considered by these models are: the round trip distance traveled between the origin (*o*) and destination (*d*), the type of fare class chosen by the traveler (first, business class and unrestricted coach class and restricted coach class), the type of airport (large hub, medium hub, small hub, or non hub), whether or not the route is served by a low cost airline and the airline market concentration between the *o-d* pair. The models suggest that competition at the destination airport is more critical than the competition at origin airport for coach class fares and vice a versa for business class fares. Models suggested in this thesis predict air fares with R-square values of 0.3 to 0.75.

ACKNOWLEDGEMENTS

On a passage through life, one reaches a stage, where, one looks for someone to emulate. At this point in time, it has to be my advisor Dr. Antonio Trani. Dr. Trani is not only an excellent teacher of engineering, but also of life. His calm composure during my entire length of stay at Virginia Tech is admirable. I have learnt much from him, both academic and otherwise, and for this I am indebted.

I thank Dr. Hojong Baik for his continuous advice, encouragement and for helping me place this broader perspective in my research.

Also, I thank all my friends and fellow colleagues, too numerous to list, who made my experience at Virginia Tech pleasant, as well as administrative staff at the Via Department of Civil Engineering, for their help.

Finally, I thank my family for their continuous encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 Introduction

Understanding travel cost for transportation modes is important in transportation system analysis.Primarily travel cost is one of the major factors that a traveler considers when he/she chooses the transportation mode for the trip. For any intercity transportation systems mode choice analysis, knowing the cost of travel by each mode is very significant. For instance, analysis of the impact of the introduction of a new mode to the existing intercity transportation system, it is imperative to know the cost of travel by that particular mode. This cost of travel will also help to determine the future trend in travels whether there is going to be congestion or more demand of a particular mode.

1.1 Objective

The National Aeronautics and Space Administration (NASA) intends to reduce inter-city travel time in the United States by one-half within 10 years and by two-thirds within 25 years, while keeping costs low and improving safety. Teaming with the Federal Aviation Administration (FAA), industry, and several universities, NASA has launched the Small Aircraft Transportation System (SATS) research program.Virginia Tech was a part of this alliance. It was assisgned the critical task of transportation system demand estimation.

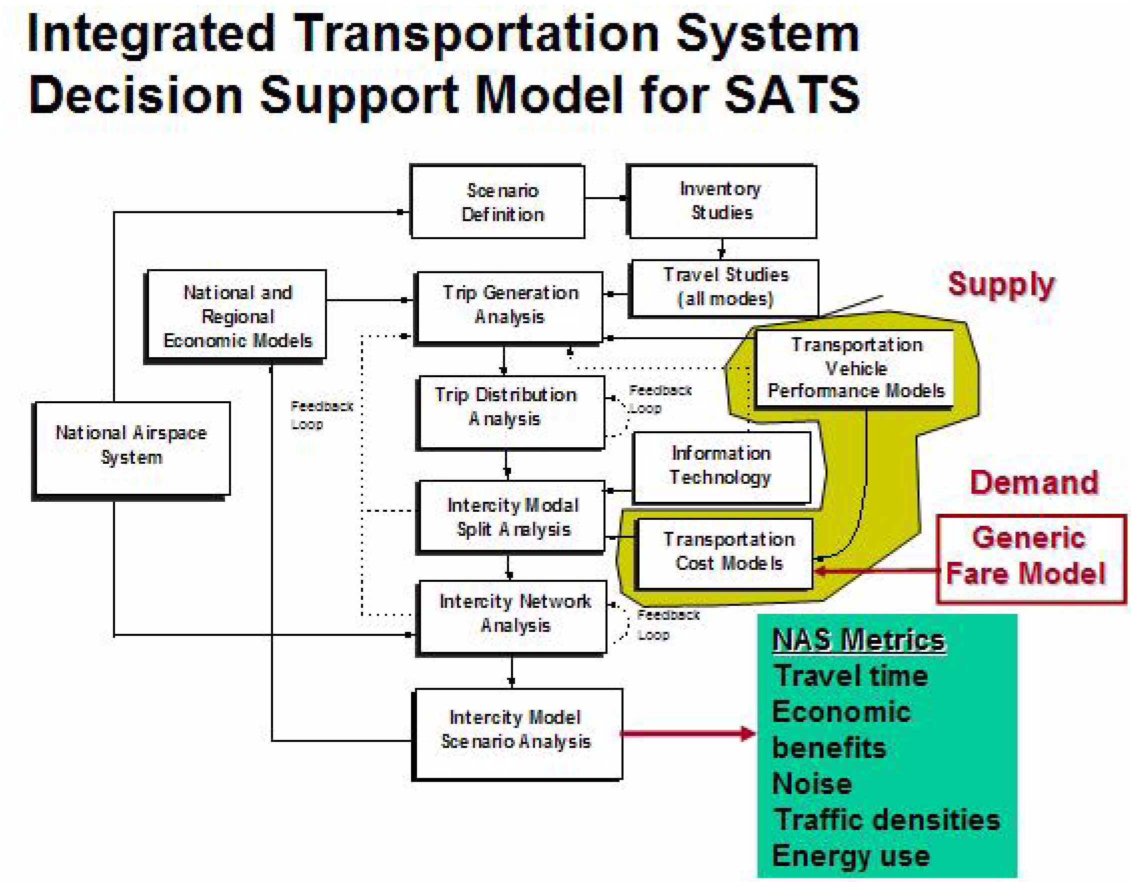


Figure 1.1 Input of Fare Model in Intercity Transportation Systems Analysis.

Transportation System demand estimation analysis on the SATS is one of the typical examples where a cost model i.e. a model that estimates travel cost, is being used as critical input in the stage of its mode choice process.This cost model is split into two sub-categories: Cost for supply side - 'Transportation Vehicle Performance Models' and Cost for demand side - 'Generic Fare Model' as shown in Figure 1.1. This thesis aims in developing this 'Generic fare model' as a demand side cost metric.The ratio of average fare to distance (fare per mile) is used as a measure of this cost of travel. Unlike other mode fares; air fare is affected by many factors and as a result it is not easy to typify the fare. The variations in cost of air travel can be attributed to many reasons. This thesis aims to understand this variation in the cost of air travel by developing series of statistical models.

1.2 Thesis Organization

The thesis first reviews different literatures and similar studies performed in the past related to airline fare. It examines the literature on the factors that affect cost of air travel and previous studies conducted in the same respect.

It then describes the different data sources that was used in the model analysis. It also explains how the data sources were modified for this study purpose.

After describing the different data sources the thesis discusses the procedure to run the entire model starting from developing a unique data set for this study purpose all the way to statistical tests performed on the model. It also describes all the input, output, SPSS syntax, and MATLAB script files used in the entire model operating procedure.

It then discusses the methodology followed in this thesis. It provides a brief explanation of the factors that affect the variability in the cost of air travel. It then develops a set of models considering all the variable factors. Finally, it gives an outline of the results of each model by providing an indication of the variables that affect the cost of air travel together with related statistics.

The thesis then concludes with suggestions and further analysis of the data.

CHAPTER 2 Literature Review

Deregulation movement has brought enormous changes to the US domestic airline industry in the past a quarter century since the Airline Deregulation Act was signed into law in 1978. It has been widely hailed as a success. The total numbers of enplanements and passenger miles have more than doubled since then, and the overall airfare has been considerably lower than it would have been had regulation continued. The traveling public in general has enjoyed a larger network of connections, greater service choices and lower fares on average. However, this overall success in deregulation should not mask the fact that all these changes in airline industry have not been even across space. There are significant geographic variations in the cost of US domestic air travel. While many people benefit from the decrease of airfares, some communities claim that they have been suffering deteriorating air travel services, mostly in the form of high airfares or short of services, namely the problem of "pockets of pain"(1). Given the important place of air travel in the whole economy, the geographic variations of air travel cost may have significant implications for regional economies and overall public welfare.

After the removal of the restrictions posted on airline industry in regulation years, airfares have taken a more and more complex structure. Traditional bond between airfares and distance has broken down as shown in Figure 2.1.

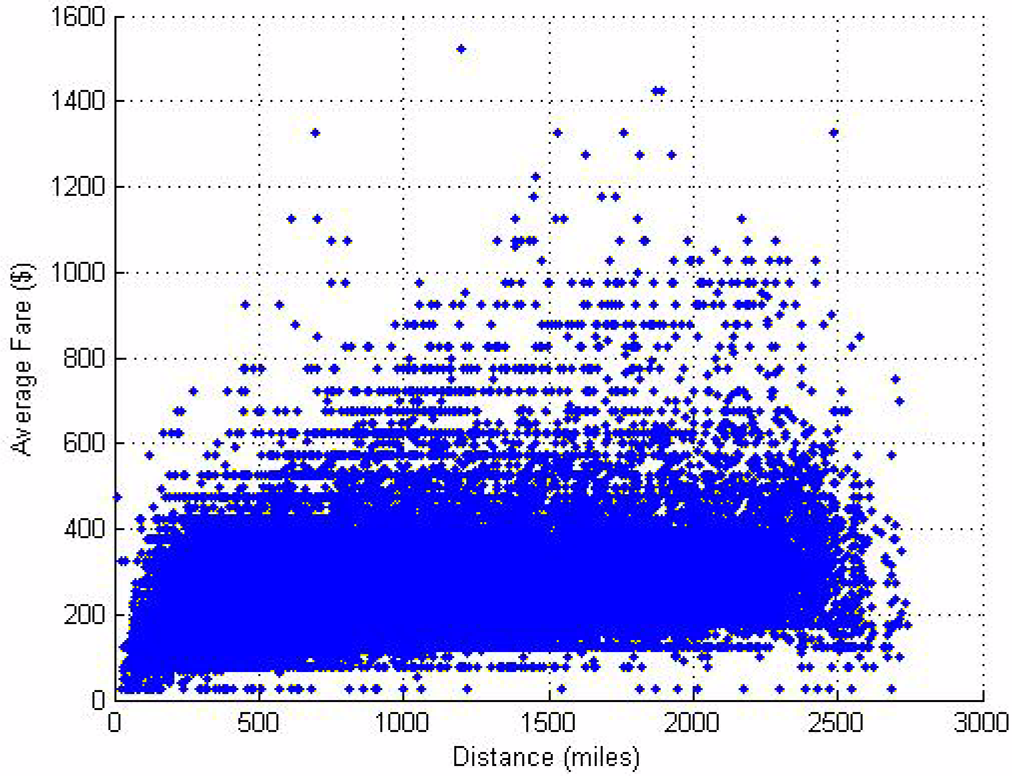


Figure 2.1 Average fare ($) vs. Distance (miles).

Instead, airfares are heavily influenced by factors such as scale economies, competition, airport congestion and airline marketing strategies. In general, longer flights tend to have lower average cost because the fixed costs associated with each flight can be spread over a longer distance. Average cost may also be expected to be lower in markets with larger passenger volume since airlines in those markets are able to use larger planes and achieve higher load factors. The level of competition is another factor that may influence air travel cost. Because airlines may compete each other in different forms, and competition can happen both at the airports and in the route markets, the influences on cost from competition have been quite complicated. Whether competition has impact over airfares had been a controversial issue even before the implementation of the airline deregulation. One theoretical foundation for deregulation was the contestable theory, and the contestability theory in its pure form suggests that the number of actual competitors should have no effect on prices. Studies about airfares have been plenty. Spatial analysis techniques, however, have been used rarely. Most of the previous studies have adopted a standard multivariate linear regression approach by regressing airfare on a series of pertinent variables and using ordinary least squares method to estimate the parameters. This thesis also develops series of multiple linear regression models to estimate average fare paid between any two destination pairs. It aims to estimate regression models that would be able to predict fares on a nationwide level. Hence, the models developed in this are generic models that are used as critical inputs as a measure of cost of travel in the mode choice process of transportation systems analysis of SATS.

2.1 Regression Models

Anderson et. al (2) used data from Domestic Airline Fares Consumer Report to conduct descriptive and statistical analyses of the spatial variability of air travel costs. The ratio of average fare to distance fare per mile was used as a general measure of relative cost, which is observed to vary significantly across cities. Figure 2.2 plots the average fare per mile against the flight distance. A clear inverse curvilinear relationship is displayed. This confirms the claim that longer trips tend to have lower fare per mile or yield because the fixed costs associated with each flight can be spread over a larger number of miles.

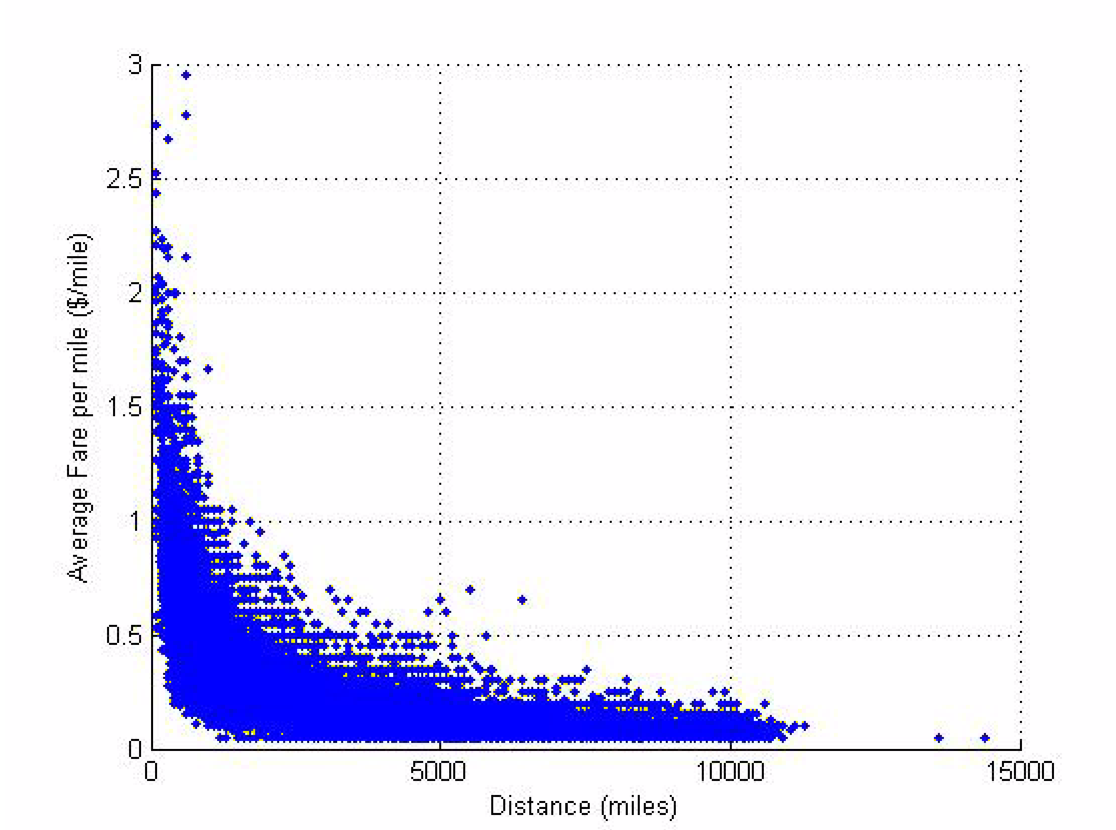


Figure 2.2 Average Fare per mile ($/mile) vs. Distance (miles).

Two regression analyses seek to identify the factors that drive this variation. The first take weighted average fares in the 50 largest cities as its dependent variable.The model is a city based regression models.

$$f_i = \beta_0 + \beta_1 D_i + \beta_2 S_i + \beta_3 P_i + \beta_5 X_i + e_{ij} \tag{2.1}$$

where,

f_i = Average fare per mile from city i;

P_i= Total Passengers Originating from city i;

D_i= Weighted Average distance from city i;

S_i= Weighted Average Market share of airline at city i;

X_i = a vector of factors other than scale, distance, and market share that may affect fares;

e_i = stochastic residual term.

The second takes average fares in more than 6,500 original destination pairs as its dependent variable. The model estimates average yield between 6,500 origin destination pairs.

$$f_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 S_{ij} + \beta_3 P_{ij} + \beta_5 OP_i + \beta_6 DP_j + e_{ij} \tag{2.2}$$

where,

f_{ij}= Average fare per miles between origin airport i and destination airport j;

P_{ij}= Total Passengers between origin airport i and destination airport j;

D_{ij}= Weighted Average distance between origin airport i and destination airport j;

S_{ij}= Weighted Average Market share of airline between origin airport i and destination airport j;

OP_i = Origin Size of the airport i;

DP_j= Destination Size of the airport j;

e_{ij} = stochastic residual term.

Results of both analyses indicate that fare per mile declines with the length of trip. They both also indicate a large and statistically significant competitive effect, whereby single-carrier market share has a positive effect on the fare. Regional variations are indicated in the results of the first regression, with higher fares in the Northeast and Midwest than in the West and South. The second set of regression results also indicates that, other things being equal, fares with origins or destinations in large cities tend to be higher. This thesis further works on these regression models by considering other explanatory variables that affect travel cost.

2.2 Herfindahl-Hirschmann Competition Index

The Herfindahl-Hirschmann Competition Index is a measure of industry concentration; a value of 1 corresponds to a monopoly; 0.5 corresponds to an industry with two equal-sized firms, 0.33 to an industry with three equal-sized firms (3). Herfindahl Index (HI) is the sum of the square of the percentages of the available seats by each airline (pa) in a particular market segment.

$$HI = p_1^2 + p_2^2 + p_3^2 + p_4^2 + ... + p_a^2 \quad (2.3)$$

where,

p_a = maket share of airline in a market segment.

a = airline under consideration.

HI = Herfindahl Index at the airport.

Herfindahl index as a measure of competetion is used as an explanatory variable to measure the cost of travel for this study purpose.Winston et. al. (4) explains the affect of deregulation on airline pricing. By using time-series analysis, the paper claims that deregulation on an average has brought down airline fares. But the decrease has not been uniform .The yield (fare/ mile) has decreased in the long-haul segment (>900 Miles) but increased in the short (<900 Mile) segment, since regulation before 1978 purposely kept the long distance fares higher. The paper uses cross-sectional time-series to show that increase in competition always brings down fares.

2.3 Low Cost Carriers

Charles Najda 's (5) analysis seeks to determine if the pricing strategies of competitors differ depending on a low-cost carrier's presence on a route. The presence of a low-cost carrier is a more important determinant of the competitiveness of a particular route than the extent of route and hub concentration on that route. The emerging significance of the low-cost carrier may indicate a shift in the structure of the airline market away from hub-and-spoke networks and towards point-to-point networks. The percentage of airline market share by low-cost carriers is significant.The presence of low-cost carriers between a market segment tends to drive down the fares significantly. This thesis considers the presence of low-cost carriers and uses this presence in the model as an explanatory variable.

2.4 Trip Length

At a national level it is common to express the amount of air travel in terms of the number of revenue passenger miles flown or the number of enplaned passengers (6). This provides a way to resolve the difficulty of how to aggregate measures of air travel in many different markets of many different distances. However, information about the distribution of trip lengths is lost in the process. This information is of interest for a number of reasons. The type of aircraft that is most appropriate for different markets depends on the distances involved. The length of the trip is also likely to influence traveler behavior in terms of the importance of convenient access to airports, frequency of service, and willingness to make intermediate stops.A related issue is how the cost of air travel varies with the length of the trip. The cost structure of airline service is such that the cost per mile flown reduces with increasing trip length, and this is reflected in typical airline fares. However, other factors also influence fares, such as airline competition, and the fare structure may or may not reflect the costs involved for trips of different lengths. Sheng-Chen et. al (6) studies distribution of trip lengths and associated average fares in the U.S. domestic air travel market.The studies determined that over half of the US markets are under 1000 miles ,and more than 60 percent of passengers have trip lengths under 1000 miles. In general, the number of passengers decreases with increasing trip length. However, a secondary peak in the distribution occurs in the range between 2200 miles and 2800 miles due to coast-to-coast operations. The thesis used this study to develop clusters in trip distance i.e. trip length to further understand the relationship between trip distance and cost of travel.

2.5 Heteroscadacity

For estimation of regression coefficients and for regression inference to be correct it is assumed that:

1. Equation is correctly specified.

2. Error Term has zero mean.

3. Error Term has constant variance.

4. Error Term is not auto correlated.

5. Explanatory variables are fixed.

6. No linear relationship between RHS variables.

When assumption 3 holds, the errors u_i in the regression equation have common variances regression model has a scalar error covariance matrix (assuming also that there is no autocorrelation), where 'scalar' is another word for constant. This is also known as homoscedasticity. When assumption 3 breaks down,the regression model has a non-scalar error covariance matrix. This is also know as heteroscedasticity.

2.5.1 Scalar Error Covariance Matrix

Assumption 4 of ordinary least sqaure (OLS) requirements states that the sampling distributions for each of the residuals are not correlated with any of the others. So, for any two observations, the residual terms are uncorrelated: $cov(u_1, u_2) = 0$; or more generally: $cov(u_i, u_j) = 0 \; V \; i, j$.

$$\operatorname{cov}(u_1, u_2, \ldots u_n) = \begin{bmatrix} \operatorname{var}(u_1) & \operatorname{cov}(u_1, u_2) & \cdots & \operatorname{cov}(u_1, u_n) \\ \operatorname{cov}(u_2, u_1) & \operatorname{var}(u_2) & & \operatorname{cov}(u_2, u_n) \\ \vdots & & \ddots & \vdots \\ \operatorname{cov}(u_n, u_1) & \operatorname{cov}(u_n, u_2) & \cdots & \operatorname{var}(u_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \quad \text{where } \sigma^2 \text{ is a scalar}$$

Figure 2.3 Covariance Matrix.

Assumption 3 states that the variance of each of the sampling distributions should be the same, so a covariance matrix of residuals from repeated samples should have a constant value ('scalar') down the diagonal and zero's off the diagonal.

2.5.2 Homoscedastic errors have a scalar error covariance matrix

Assumption 5 of OLS requirements states that regressors are fixed. This means that, as in an experiment, the regressors (or control variables) can be repeated. For each value of the control variable, the scientist will observe a particular effect (i.e. a particular value of the dependent variable). In repeated experiments,he/ she can keep the values of the control variables the same, and observe the effects on the dependent variable. There will thus be a range of values of y for each controlled and repeatable value of x. If we plot observed values of y for given values of x repeated samples, then the regression line will run through the mean of each of these conditional distributions of y. This is shown in Figure 2.4.

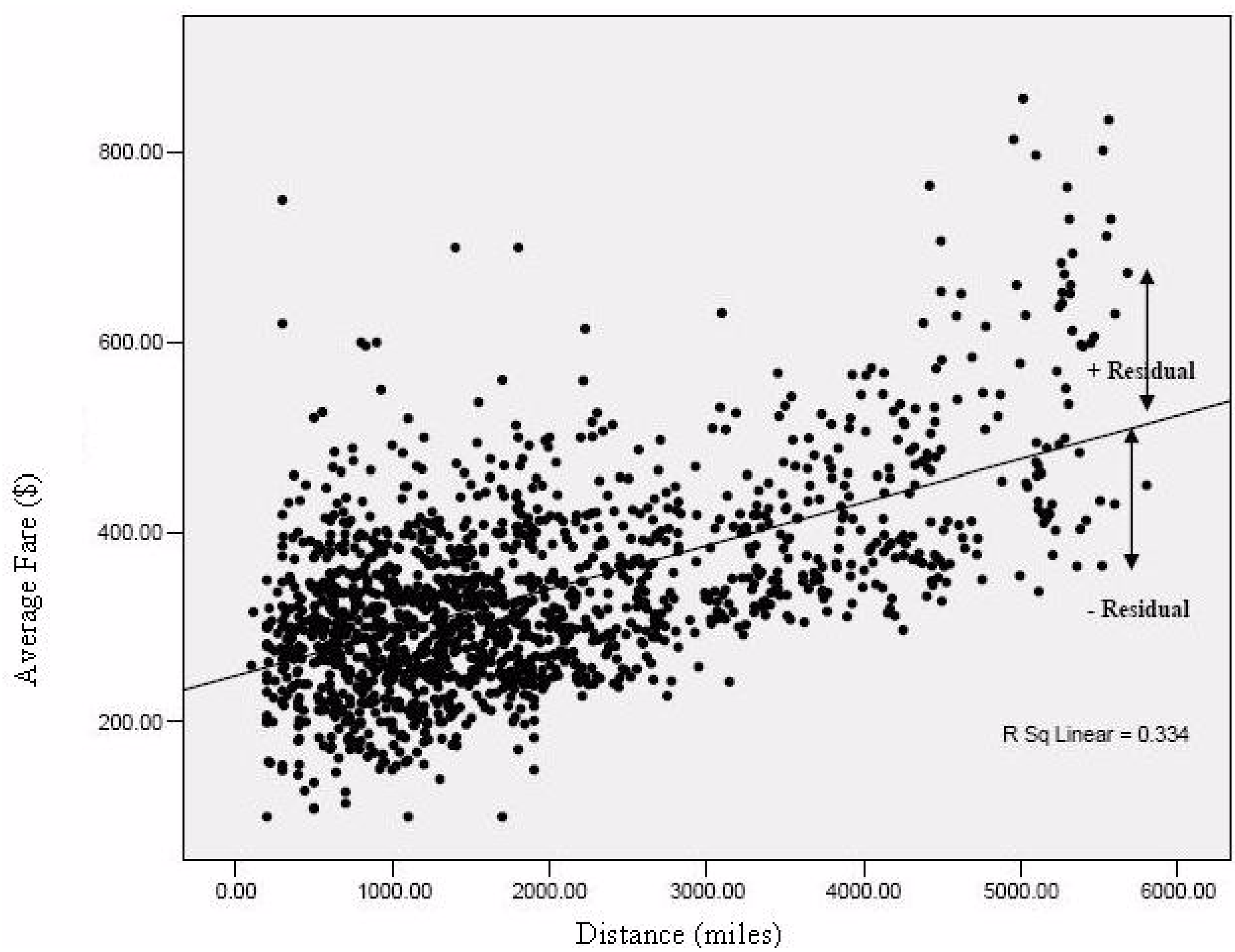


Figure 2.4 Simple Linear Regression.

However, each time a regression is run, it is run on a particular sample, for which there may only be one value of y for a given x or many values, depending on the experiment. As such, for each sample, there will be a slightly different line of best fit and estimates of a and b (the intercept and slope coefficients) will vary from sample to sample. The variability of b across samples is measured by the standard error of b, which is an estimate of the variation of b across regressions run on repeated samples. Standard Error of b can be estimated from within the current sample because the variability of the slope parameter estimate will be linked to the variability of the y-values about the hypothesized line of best fit within the current sample. In particular, it is likely that the greater the variability of y for each given value of x, the greater the variability of estimates of a and b in repeated samples . Hence, from the variability of y for a given value of x it is easy to provide an estimate of the sampling variability of b.Recall that the value of the residual for each observation i is the vertical distance between the observed value of the dependent variable and the predicted value of the dependent variable (i.e. the difference between the observed value of the dependent variable and the line of best fit value). Assume that Figure 2.4 is a plot from a single sample. Each one of the residuals has a sampling distribution, each of which should have the same variance -- “homoscedasticity”. Clearly, this is not the case within in this example, and so is unlikely to be true across samples. Although the sampling distribution of a residual cannot be estimated precisely from within one sample (by definition, one would need to run the same regression on repeated samples) as with standar error of b, one can get an idea of how it might vary between samples by looking at how it varies within the current sample.

2.5.3 Causes

There may be multiple reason for the variance of the residuals to change over the course of the sample. The error term may be correlated with either the dependent variable and/or the explanatory variables in the model, or some combination (linear or non-linear) of all variables in the model or those that should be in the model.

2.5.4 Detection: Whites Test

The most general test of heteroscedasticity is White’s test (7) .Procedure for White’s test:

Step 1: Run an OLS regression – use the OLS regression to calculate uhat 2 (i.e.square of residual).

Step 2: Use uhat 2 as the dependent variable in another regression, in which the regressors are: (a) all “k” original independent variables, and (b) the square of each independent variable, (excluding dummy variables), and all 2-way interactions (or crossproducts) between the independent variables. The square of a dummy variable is excluded because it will be perfectly correlated with the dummy variable. Call the total number of regressors (not including the constant term) in this second equation, P.

Step 3: From results of equation 2, calculate the test statistic: $nR^2 \sim X^2 P$

where n = sample size, and R^2 = unadjusted coefficient of determination. The statistic is asymptotically ,i.e. in large samples, distributed as chi-squared with P degrees of freedom, where P is the number of regressors in the regression, not including the constant.

2.5.5 Notes on White's test

The White test does not make any assumptions about the particular form of heteroskedasticity, and so is quite general in application. It does not require that the error terms be normally distributed. However, rejecting the null may be an indication of model specification error, as well as or instead of heteroskedasticity. Generality is both a virtue and a shortcoming. It might reveal heteroscedasticity, but it might also simply be rejected as a result of missing variables. It is "nonconstructive" in the sense that its rejection does not provide any clear indication of how to proceed. However, if one uses White's standard errors, eradicating the heteroscedasticity is less important.

CHAPTER 3 Data Sources

Over the years, government agencies, airline industry officials and survey research firms have amassed data that represent the collective experience of travelling public and the airline industry.This statistical information is used in two ways in this study. In this chapter these data sources are first interpreted and then modified for this study purpose.In later chapters this modified data provides the basis for statistical models of travel cost behavior (called Fare Models hereon).

3.1 Data Sources for the Fare Model

There are several data sources are available they are as follows:

A) *Sabre AirPrice* - data provided by SABRE.

B) *Airline Origin and Destination Survey (DB1B)*-data provided by the Bureau of Transportation Statistics.

A. Sabre AirPrice

The Sabre AirPrice is the most comprehensive data available for fares. It has databases which enables retrieving both published and unpublished fares by origin and destination pairs. The main advantage of the databases is that it has a historical data of fares which enables comparison of prices over time i.e gives an idea of the supply and demand between an O-D pair, which is not available for general public since historical data of fares would enable us to understand the airline pricing startergies and the 'crazy math of airline ticket pricing'. However since such a data is only available online it makes the availability of the data biased to only online users and the databases are available at a hefty price from SABRE.

B. Airline Origin and Destination Survey (DB1B)

The main data source for our analysis is Airline Origin and Destination Survey(DB1B) provided by the Bureau of Transportation Statistics. The DB1B is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. This database is used to determine air traffic patterns,air carrier market shares, and passenger,freight,and mail cargo flow. The survey covers a sample of airline tickets from reporting carriers beginning from 1994 to 2004 over a quarterly period.

3.2 Explanation of DB1B

Three different categories of table are available in DB1B:

1) DB1BCoupon: This table provides coupon specific information for each domestic itinerary of the Origin and Destination Survey,such as operating carrier ,origin and destination airports,number of passengers,fare class,coupon type,trip break indicator and distance.Appendix A gives information about this table.

2) DB1BMarket: This table contains directional market characteristics of each domestic itinerary of the Origin and Destination Survey,such as reporting carrier,origin and destination airports,prorated market fare,number of market coupons,market miles flown,and carrier change indicators.Appendix A gives information about this table.

3) DB1BTicket: This table contains summary characteristics of each domestic itinerary of the

Origin and Destination Survey,including the reporting carrier,prorated market fare,number of market coupons,market miles flown,and carrier change indicators.Appendix A gives information about this table.

3.3 Comparison of the Data Tables

Although the three databases are a source of extensive information, they were merged to create a unique data source for this study purpose.Inorder to understand the relationship between the three databases consider the Table 3.1 given below.

Table 3.1. Information about Data Sets in each Table.

	TABLE TYPE		
FIELD	COUPON	MARKET	TICKET
Itinerary ID	Y	Y	Y
Market ID	Y	Y	
Year	Y	Y	Y
Quarter	Y	Y	Y
Sequence No.	Y		
Coupons	Y	Y	Y
No.of Passengers	Y	Y	Y
Fare Class	Y		
Market Fare		Y	
Roundtrip Fare			Y
Distance	Y	Y	Y
Origin	Y	Y	Y
Destination	Y	Y	Y
Carrier	Y	Y	Y
Itinerary Yield			Y

This table compares the different unique and common fields associated with the databases Coupon, Market and Ticket. The three databases have several common fields that associate the databases with each other. One of them is Itinerary ID that was used as a unique ID to match or combine the three different databases. The field Market ID is common for Coupon and Market database. This ID was used to merge Coupon and Market database.The field Fare Class is unique to Coupon, whereas Market Fare is unique to Market and finally Round-trip Fare is unique to Ticket. Combining the three tables links the Market & Round-trip fare with its corresponding Fare Class.The Ticket database provided only origin airport information for an itinerary. This combined database identified destination airports for individual itinerary in the Ticket database.This combined data table or the modified database is one of the inputs used in Fare Models.Table 3.2 shows an illustration of the combined database. SPSS 12.0 was used as the primary database tool for the analysis.

Table 3.2. Combined Tables - Coupon, Market and Ticket.

	TABLE TYPE		
FIELD	COUPON	MARKET	TICKET
Itinerary ID	Y	Y	Y
Market ID	Y	Y	
Year	Y	Y	Y
Quarter	Y	Y	Y
Sequence No.	Y		
Coupons	Y	Y	Y
No.of Passengers	Y	Y	Y
Fare Class	Y		
Market Fare		Y	
Roundtrip Fare			Y
Distance	Y	Y	Y
Origin	Y	Y	Y
Destination	Y	Y	Y
Carrier	Y	Y	Y
Itinerary Yield			Y

FIELD	Combined Coupon, Market & Ticket
Itinerary ID	Y
Market ID	Y
Year	Y
Quarter	Y
Sequence No.	Y
Coupons	Y
No.of Passengers	Y
Fare Class	Y
Market Fare	Y
Roundtrip Fare	Y
Distance	Y
Origin	Y
Destination	Y
Carrier	Y
Itinerary Yield	Y

3.4 Interpretation of the Data Sets: Coupon, Market, and Ticket

In order to understand the relationship between Coupon, Market and Ticket databases a round-trip flight from Albuquerque (ABQ), NM to Atlanta (ATL), GA via Dallas (DFW), TX is considered. This itinerary has a unique Itinerary ID - 200019 in database as shown in Figure 3.1 and was used as a key to merge the three databases. Figure 3.2 is an illustration of this example using a graphical map.

DB1BCoupon Database

break	dest	des	distance	fare	itingeot	itinid	mktid	opcar	origin	origin	passenge	rpc	seqnum	tkc
X	ATL	GA	1269	X	2	200019	2000114	AA	ABQ	NM	1.00	AA	1	AA
NULL	DFW	TX	732	X	2	200019	2000115	AA	ATL	GA	1.00	AA	2	AA
X	ABQ	NM	569	X	2	200019	2000115	AA	DFW	TX	1.00	AA	3	AA

DB1BMarket Table

airportg	bulkfar	dest	deststat	itingeot	itinid	v15	mktfare	mktid	mktmile	nonstop	opcarri	origin	origin	pas	rpc	tkcarr	v37
ABQ:ATL	0	ATL	GA	2	200019	3	98.00	2000114	1269	1269	AA	ABQ	NM	1	AA	AA	AA
ATL:DFW:ABQ	0	ABQ	NM	2	200019	3	101.00	2000115	1301	1269	AA	ATL	GA	1	AA	AA	AA:AA

DB1BTicket Table

ar	coupons	distance	v4	dollarcr	itinfare	itingeot	itinid	itinyiel	online	origin	origina	origin	passen	quarter	rpcarri	roundtr
0	3	2570	6	1	199	2	200019	.08	1	ABQ	1	NM	1	1	AA	1

Combined Coupon, Market, and Ticket

ecla	itinid	airportg	dest	nonstop	v23	origin	itinfare	itinyiel	milesfl	online	ori_t	passeng	rpcarrie	roundtri
	200019	ABQ:ATL	ATL	1269	AA	ABQ	199	.08	2570	1	ABQ	1	AA	1

Figure 3.1 Methodology to Combine Coupon, Market, and Ticket.

The Itinerary ID is same for all the legs where as the Market ID is different for the first leg and the last two legs. Each leg is designated by a sequence number. The total number of Coupons is equal to the total number of legs for an itinerary.Hence the above Itinerary Id has three coupons. The round-trip fare information in the Ticket database is the sum of the market fares from the Market database. This round-trip fare along with the origin and destination airport Id and passenger flow information are used to develop plethora of different origin and destination tables of mean fares, passenger flows etc. that were used to develop the Fare model.

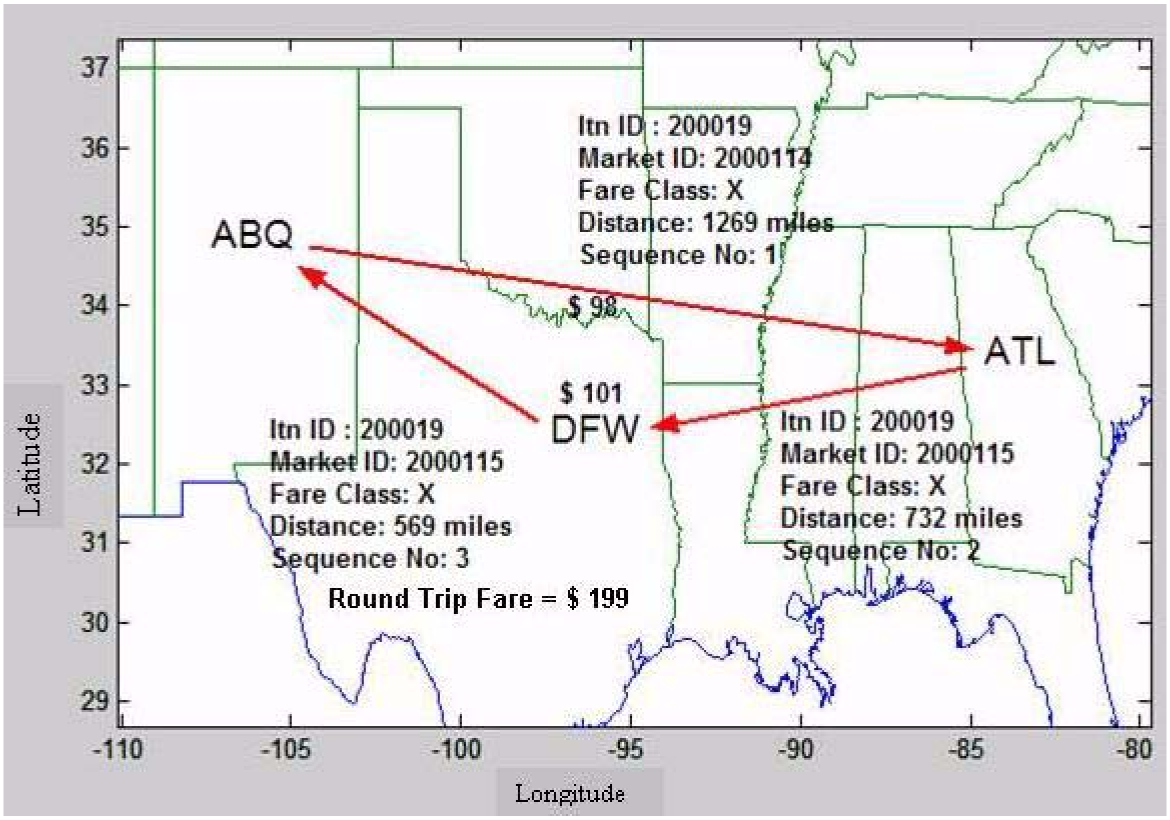


Figure 3.2 Visual Interpretation of Coupon and Market Database.

3.5 Data Filtering

Most large commercial databases have erroneous data. These data elements may lead to outliers and or inaccurate results during the statistical analysis. Hence, these outliers have to be deleted from the database before performing the model analysis. The following data filter is applied to the databases:

1) All Zero and Null value of fares were deleted.

2) All round trip fares having values less than 50 $ more than 5000$ were deleted.

3) All round trip distance less than 200 statue miles and more than 10,000 statue miles were deleted.

Appendix A has the SPSS syntax that automates the task of combining the two databases and deletes outliers and inconsistencies in the data sets.

3.6 Fare Class

The different fare classes and their corresponding codes that are considered for fare classification in DB1B are as follows:

1.C - Unrestricted Business Class

2.D - Restricted Business Class

3.F - Unrestricted First Class

4.G - Restricted First Class

5.U - Unknown

6.X - Restricted Coach Class

7.Y - Unrestricted Coach Class

Passenger flow statistics by fare class for the year 2000 indicate that nearly 72% of the passengers in National Airspace System travel by restricted coach class (X). This is indicated in Table 3.3.

Table 3.3. Distribution of Fare Classes for year 2000.

Fare Class Code	Fare Class	Passengers	Percentage
C	Unrestricted Business Class	891,830	0.15%
D	Restricted Business Class	880,730	0.15%
F	Unrestricted First Class	34,905,910	5.88%
G	Restricted First Class	52,902,340	8.91%
U	Unknown	3,204,860	0.54%
X	Restricted Coach Class	425,622,430	71.71%
Y	Unrestricted Coach Class	66,689,530	11.24%
	Unknown	8,407,890	1.42%
	Total	593,505,520	100.00%

3.7 Official Airline Guide (OAG)

The OAG contains information about flight schedules and frequencies for airlines throughout the world. The OAG includes information like:

(i) Air Carrier- Indicates the airline that operates the service between the airport pairs.

(ii) Origin Airport - The originating point of the flight.

(iii) Destination Airport - The endpoint of the flight.

(iv) Airtime - Time spent in air.

(v) Ground time - Time spent on the ground including waiting and transfer times at airports.

(vi) Seats - Number of available seats in the flight.

(vii) Distance - Flight distance in miles.

(viii) Frequency - Number of flights between the origin and destination during the considered time intreval.

(ix) Aircraft type - The name and model of aircraft that was flown between the origin and destination airports.

The OAG is used primarily in this study to compute Herfindhal competition index and to identify the presence of low-cost carrier services between origin and destination pairs.The OAG was used along with the DB1B database to determine the mobility for air travelers taking commercial flights. The combination of these two databases can also yield other important information about the structure of the airline network of the US such as location of hubs, percentage of direct flights and load factors for different markets.

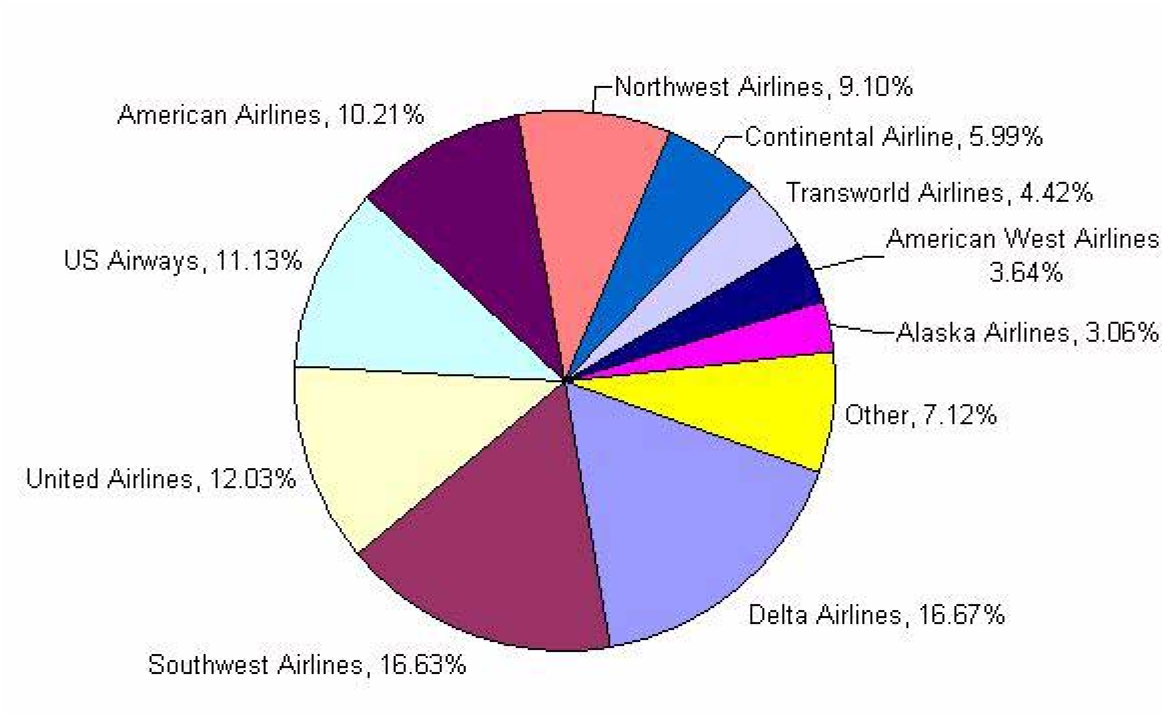


Figure 3.3 Airline Market Share for the Year 2000 (Data Source: OAG).

3.8 Model Flow

The model used a combination of different data sources to develop a unique statistically significant and robust dataset for the Fare Model. It combines OAG and DB1B to understand the supply and demand side of cost of air travel in NAS. The data set is a comprehensive data source that could be used to determine:

1)Average fare ($) between origin and destination pair.

2)Passenger flows between origin and destination pair.

3)Competition index in a market segment.

4)Low-Cost presence in a market segment.

The over all model flow for the fare model is shown in Figure 3.4.

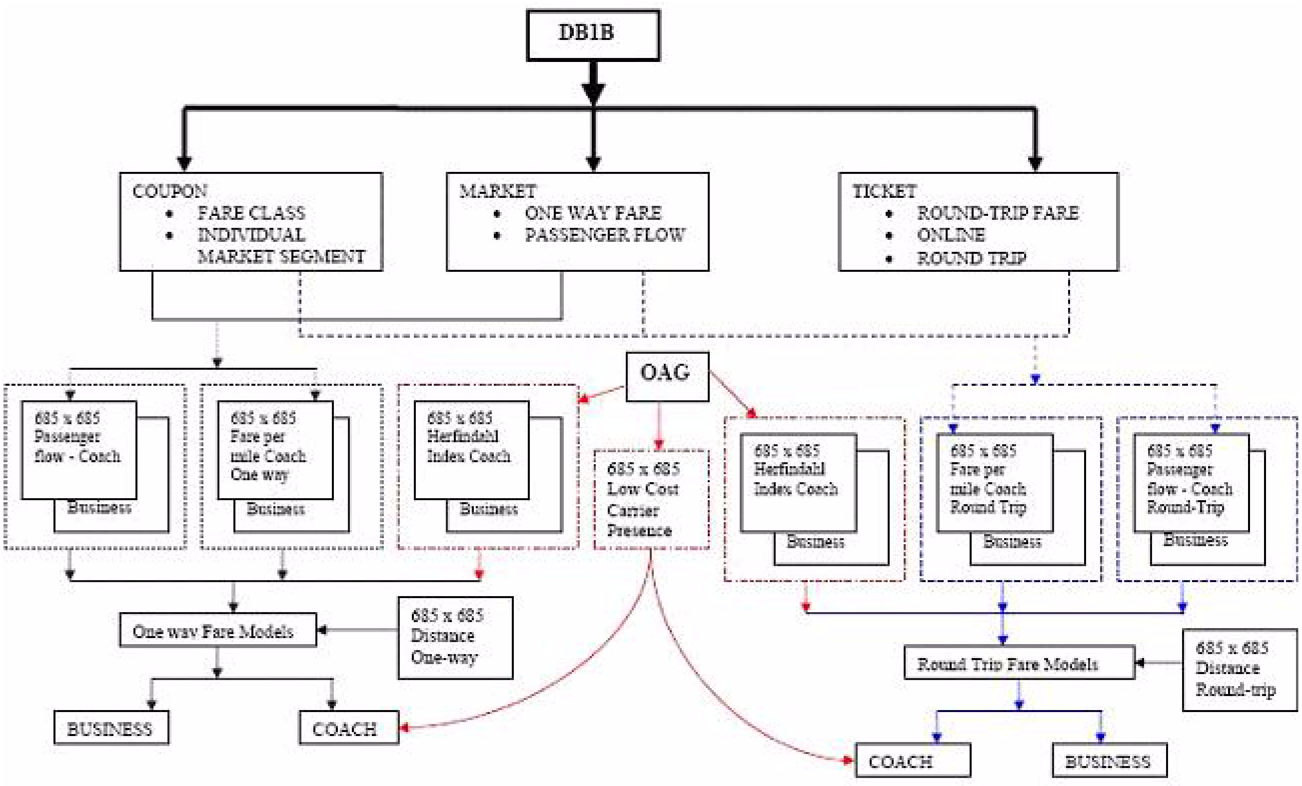


Figure 3.4 Model Flow for Fare Model.

CHAPTER 4 Data Analysis Procedure

This chapter describes the model operating procedure associated with the Fare Model. The chapter first explains in brief the different software's used in the model. It then illustrates in detail the model operating procedure via a Flow Chart.Finally it describes the input files, Matlab code or SPSS syntax and output files used and generated , using the Flow Chart ,during Fare Model process.

4.1 Software and Programs

1) **MATLAB**: MATLAB refers to both the numerical computing environment and to its core programming language. Created by The MathWorks, MATLAB allows easy matrix manipulation, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs in other languages.In this thesis, MATLAB was used primarily as a programming tool.

2) SPSS: SPSS (originally, Statistical Package for the Social Sciences) was released in its first version in the 1960s, and is among the most widely used programs for statistical analysis in social science. It is also used by market researchers, health researchers, survey companies, government, education researchers, and others. The SPSS Data Editor is useful for performing statistical tests, such as testing for correlation, multicollinearity, and hypotheses; it can also provide a researcher with frequency counts, sort data, rearrange data, and serve as a useful database tool.The SPSS syntax editor is useful for automating statistical and database procedures. It also helps to increase productivity with less manual intervention required to perform data analysis.In this thesis, SPSS was used primarily as a database and statistical tool.

3) Curve Expert: Curve Expert is a comprehensive curve fitting system developed for Windows environment. X & Y data can be modeled using a toolbox of linear regression models, nonlinear regression models, interpolation, or splines. Over 30 models are built-in, but custom regression models may also be defined by the user. Full-featured graphing capability allows thorough examination of the curve fit. The process of finding the best fit can be automated by letting Curve Expert compare the data to each model to choose the best curve. This program was designed to be simple but powerful, so that all users can obtain a model for their data quickly and easily.In this thesis, Curve Expert was used primarily as curve fitting tool to estimate Fare models also known as Harris Models.

4.1.1 File type & Associated File extensions

Before getting into the model operating procedure, it is imperative to know about the details of some of the different file types and associated file extensions that are required to run the model. The file types used in the model are:

- Matlab mat file: MAT-files are double-precision, binary, MATLAB format files (e.g.fare_roundtrip_bus_2000.mat).
- Matlab m file: M-files are ordinary text files that are created using a text editor to write functions and programs in MATLAB(e.g.fare_twoway_oneway.m).
- SPSS sav file: Sav-files are SPSS data files (e.g. od_cmt_2004.sav).
- SPSS sys file: Sys-files are SPSS syntax files (e.g. Syntax_2004.sys).

- DAT file: DAT -files are tab limited data files(e.g. od_cmt_2004.dat).

4.2 Associated Input Files, Associated Code or Syntax files & Generated Output files

The overall model flow for the Fare Model is as shown in Figure 4.1. The left hand side of the model flow indicates the different softwares used to perform the requiste model analysis.The model flow is split into the following critical steps:

STEP 1) DB1B data manipulation & modification: This step involved developing a new modified data set used as an input in the Fare Model analysis. The main software used during this step was SPSS. This step involved merging three distinct DB1B data sets i.e. Coupon, Market & Ticket. Chapter 3 explains the process of combing these three data sets using SPSS data and syntax editor. '**Syntax_2004.sps**' was the syntax written in SPSS to automate the task of merging the above three databases.

STEP 2) Computation of mean fare, distance & passenger flows: This step of the Fare model involved computing origin - destination (o-d) tables for: average fare in dollars between an o-d pair, average flying distance in statue miles between an o-d pair, and total passenger flows between an o-d pair using the database developed in step one. This entire procedure of computing the o-d tables was performed by MATLAB program or code '**fare_oneway_twoway.m**'.

STEP 3) Generation of Input data for fare models: This step involved generating input data for developing the fare models i.e. Harris & Multiple linear regression models. The MATLAB o-d matrices developed in step two were converted into tab limited data tables by MATLAB code 'determine_curve_fit.m' and 'multiple_reg_data.m'. The outputs from these programs were used as an input for Harris & multiple linear regression models respectively.

STEP 4) Fare Model development: This final step involved developing the fare models using SPSS & Curve Expert 1.3. Statistical tests were performed on models using SPSS to test for model integrity and accuracy.

4.2.1 Associated Input Files

The different input files used during the Fare model computation are as follows:

- **DB1B:** The DB1B is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. This database is used to determine air traffic patterns, air carrier market shares, and passenger, freight, and mail cargo flow. The survey covers a sample of airline tickets from reporting carriers beginning from 1994 to 2004 over a quarterly period. Coupon, Market and Ticket are the three different databases available. These databases are used as an input in Syntax_2004.sps.
- **airports_oag.mat:** List of 690 airports obtained from FAA and arranged in an ascending order.
- **airlines_db1b.mat:** List of all the airlines from DB1B arranged in an ascending order.
- **find_airports_all.m:** MATLAB function that determines the location of queried airport from airports_oag.mat list.
- **find_airlines_all.m**: MATLAB function that determines the location of queried airline from airlines_db1b.mat list.
- **origin_lowcost_airports.mat**: List of airports with low-cost carrier presence.
- **lowcost_carrier_od_matrix.m**: This source code computes o-d table of average fares for o-d pairs with low-cost carrier presence.
- **lowcost_od_2000.mat**: O-D matrix with average fares for o-d pairs with low-cost carrier presence.
- **airports_690.mat**: This MATLAB structured file provides information for airport 3-digit code, latitude, longitude, and airport type associated with each airport.
- **interAirportDist_690.m**: The source code that computes the great circle distance in statue miles between given origin and destination pair.
- **interAirportDist_690.mat**: Great Circle Distance o-d matrix in statue miles.

- **herf_index.mat**: List of Herfindahl Hirschman competition index for all the airports.

4.2.2 Program Generated Output files

The different input files used during the Fare model computation are as follows:

- **fare_roundtrip_bus_2000.mat**: O-D matrix of average round trip business fare in dollars.
- **fare_roundtrip_nonbus_2000.mat**: O-D matrix of average round trip non business fare in dollars.
- **pax_rt_b_2000.mat**: Round trip passenger flow o-d matrix for business class.
- **pax_rt_nb_2000.mat**: Round trip passenger flow o-d matrix for non business class.
- **dist_roundtrip_bus_2000.mat**: Average round trip flying distance o-d matrix in statue miles for business class.
- **dist_roundtrip_nonbus_2000.mat**: Average round trip flying distance o-d matrix in statue miles for non-business class.
- **fare_pax_hindex_nb_lc_2000.dat**: This output data file from '**multiple_reg_data.m**' has specific information of average fare between o-d pair, average round trip distance in miles for the o-d pair, total passenger flow between the o-d, competition index at the origin airport, presence of low cost carriers in the market segment, and origin and destination airports types for non business class. This file is also saved as '**fare_pax_hindex_nb_lc_2000.sav**' in SPSS format.
- **fare_pax_hindex_b_2000.dat**: This output data file from 'multiple_reg_data.m' has specific information of average fare between o-d pair, average round trip distance in miles for the o-d pair, total passenger flow between the o-d, competition index at the origin airport, and origin and destination airports types for business class.
- **fpm_dist_nb.dat**: This data was used as an input to develop 'Harris Model' for the non-business class. The model was average fare per mile for non business class as

dependent variable and average round-trip distance in statue mile as independentvariable. This is output data file from '**determine_curve_fit.m**'.

- **fpm_dist_b.dat:** This data was used as an input to develop 'Harris Model' for the business class. The model was average fare per mile for business class as dependent variable and average round-trip distance in statue mile as independent variable. This is output data file from '**determine_curve_fit.m**'.
- **fpm_gcd_nb.dat:** This data was used as an input to develop 'Harris Model' for the non business class. The model was average fare per mile for non business class as dependent variable and great circle distance in statue mile as independent variable. This is output data file from '**determine_curve_fit.m**'.

The various MATLAB source code , SPSS syntax, input files used in the Fare model are listed and shown in the Appendix.The overall data flow for the model is shown in Figure 4.1.

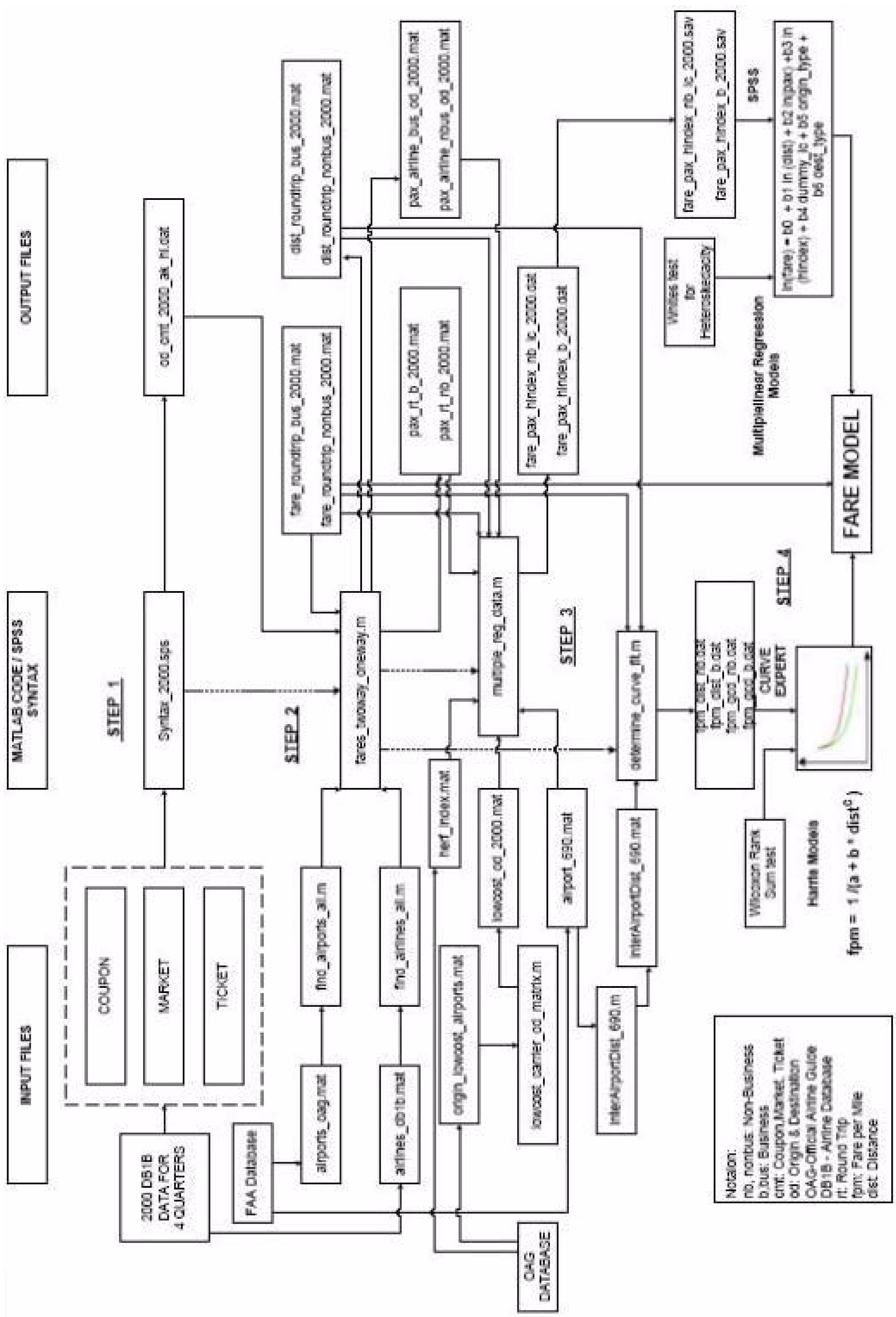


Figure 4.1 Data Flow for Fare Flow Model.

4.3 Running the Model

The file and folder structure for Fare model is shown in Figure 4.2.

Figure 4.2 File and Directory Structure for Fare model.

Once all of the files have been installed in the correct folders, the model can be run. The process for running the model is as follows:

- Start a SPSS production facility session.Run the syntax named 'Syntax_2004.sps' placed under DB1B folder.Copy the output 'od_cmt_2004_ak_hi.dat' from this session and place it under folder STEP 2.
- Start a MATLAB session. Run the source code 'fare_twoway_oneway.m'.Copy the and place outputs fare_roundtrip_bus_2000.mat, fare_roundtrip_nonbus_2000.mat, pax_rt_b_2000.mat,pax_rt_nb_2000.mat,

dist_roundtrip_bus_2000.mat, and dist_roundtrip_nonbus_2000.mat under folder STEP 3.

- Run source codes multiple_reg_data.m & data_curve_fit.m. Copy the output files fpm_pax_hindex_nb_lc_2000.dat, fpm_pax_hindex_b_2000.dat, fpm_dist_nb.dat, fpm_dist_b.dat,fpm_gcd_nb.dat & fpm_gcd_b.dat and place it under STEP 4.
- Start Curve Expert session and develop Harris Models.
- Start SPSS session and develop multiple linear regression models.

CHAPTER 5 Model Methodology and Results

This section describes the procedure used to estimate the Fare Model. Two sets of models were developed to estimate the cost of travel in US. First,a non-linear model estimates the relationship average round-trip fare and yield. The second set of models are multiple regression models that try to understand the causal relationship between average fare between any origin and destination pair and other defined explanatory variables.The models do not take into consideration the seasonal variations in the cost of travel. The main aim of the thesis was to develop generic fare models that can be effectively used to estimate the cost of travel on a national level and can be used as a cost metric during the mode choice process of SATS.The models were estimated for the base year - 2000 using the data sources explained in Chapter 3.

5.1 Methodology

For this analysis several regression models were developed using various factors to predict the variability in the fare paid. The ratio of average fare to distance, i.e. fare per mile, was used to understand the variability in the cost of air travel between an OD airport pair. A list of 685 commercial airports classified by the FAA was used in the analysis.Appendix A has this list of airports. These airports were clustered and grouped in four separate categories of airports. These airports were classified based on the total number of enplanements as shown in Table 5.1.

Table 5.1. Airport Type Classification and Code.

Airport Type	Code
Large Hub	1
Medium Hub	2
Small Hub	3
Non Hub	4

Nearly 95% of these enplanements in National Airspace System go through the Large and Medium Hubs.

5.1.1 Great Circle Distance

A great circle is the shortest path between two points along the surface of a sphere. The precise definition of a great circle is the intersection of the surface with a plane passing through the center of the planet. Thus, great circles always bisect the sphere. The equator and all meridians are great circles. All great circles other than these do not have a constant azimuth, the spherical analog of slope; they cross successive meridians at different angles. That great circles are the shortest path between points is not always apparent from maps, because very few map projections represent arbitrary great circles as straight lines. Because they define paths that minimize distance between two (or three) points, great circles are examples of geodesics. In general, a geodesic is the straightest possible path constrained to lie on a curved surface, independent of the choice of a coordinate system. A 685 x 685 table of great circle distances between the OD airport pairs from the airport list in statue mile was calculated.

5.1.2 Determination of Fare Class Category

The fare paid depends on the type of ticket purchased (i.e., restricted or non-restricted), advanced time of purchase, the OD pair demand, the airline supply cost, etc. (DOT, 2000). In this analysis, the fares were grouped into two types: First and Business class and Non Business class. Recall from Chapter 3 the list of different fare classes considered in DB1B as shown in Table 5.2.

Table 5.2. Fare Class Categories.

Fare Class Code	Fare Class
C	Unrestricted Business Class
D	Restricted Business Class
F	Unrestricted First Class
G	Restricted First Class
U	Unknown
X	Restricted Coach Class
Y	Unrestricted Coach Class

In order to determine the proper fare class category to be used for the analysis a set of fare class groups was created using the fare class categories. They are as follows:

a) Business Class - Unrestricted First Class (F), Restricted First Class (G), Unrestricted Business Class (C) and Restricted Business Class (D).

b) Coach Class - Unrestricted Coach Class (Y) and Restricted Coach Class (X).

c) Restricted Coach Class (X).

d) Unrestricted Coach Class (Y).

Using these categories as a basis for class determination;non-linear regression models were generated using the distance traveled as an independent variable. These models are plotted in Figure 5.1.

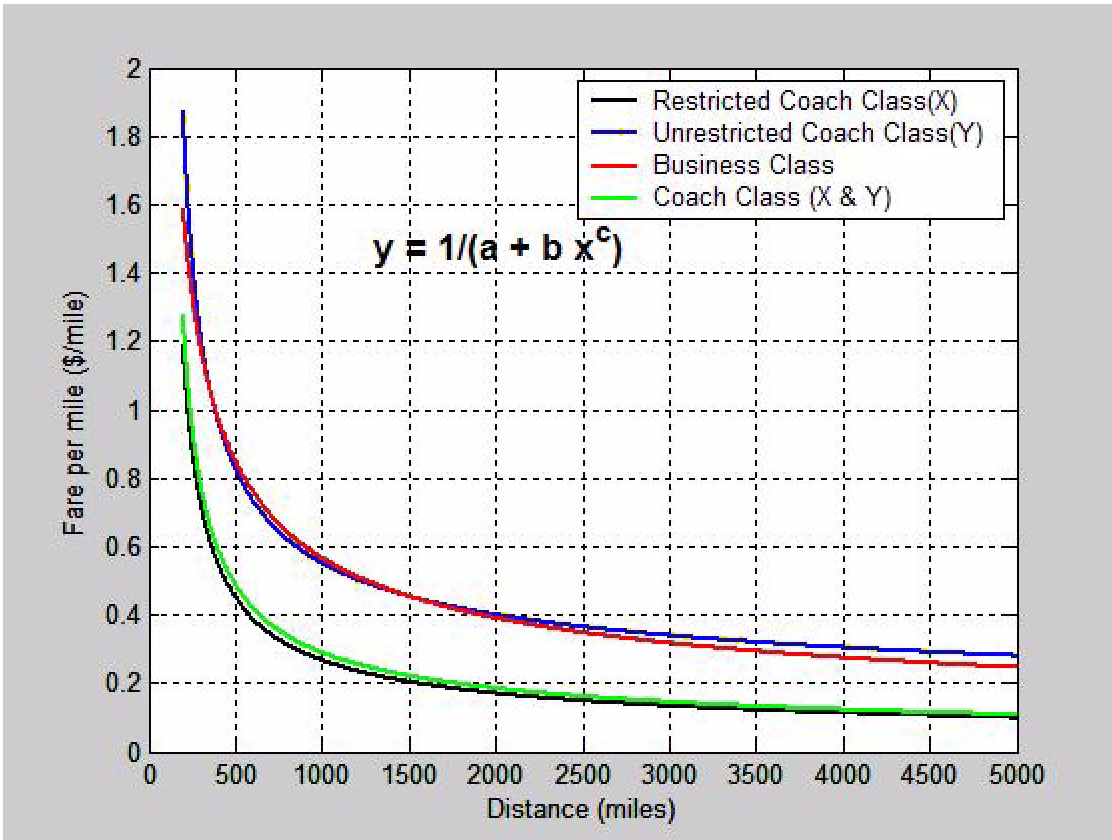


Figure 5.1 Determination of Fare Class Categories.

From Figure 5.1, Fare model for Unrestricted Coach Class(Y) behaves similar to the Business Class fare model. Hence, Unrestricted Coach Class fares were combined with Business class fares for the analysis. The final cluster of fare class groups that were used to develop the models is given below:

a) Business Class - Unrestricted First Class (F), Restricted First Class (G), Unrestricted Business Class (C) Restricted Business Class (D) and Unrestricted Coach Class (Y).

b) Non-Business Class - Restricted Coach Class (X).

5.1.3 Model Variables

1) Round Trip Distance:

After the deregulation period the relationship between airfares and distance has broken down (1). Hence a set of other factors that affect airfares is described below.

2) Market Concentration:

Competition between airlines has an important impact on the cost of air travel (1). In order to understand this competition between an o-d airport pair, the percentage of total number of seats offered by each carrier is calculated.

The total number of seats offered by each airline(a) is the sum of the product of the frequency of the type of aircraft (b) used for a flight and the number of seats in that particular type of aircraft (b) as shown in equation 1. These seats are further classified as coach and business seats. Data for airline frequency and seats is obtained from the Official Airline Guide (OAG) (6). The percentage of number of seats (pa) offered by each airline (a) for a given origin (i) and destination (j) airports pair is calculated as follows:

$$t_a = \sum_b (f_{ab} \cdot s_b) \tag{5.1}$$

$$s_a = \sum_a t_a \tag{5.2}$$

$$p_a = t_a / s_a \tag{5.3}$$

Where,

a: total number of airlines at origin (i) airport.

b: types of aircraft by each airline from origin (i) airport.

t_a: total number of seats for each airline a from origin (i) airport.

f_{ab}: frequency of aircraft type b offered by each airline a,

s_b: number of seats for aircraft type b,

s_a: total number of seats offered by an airline a from origin (i) airport.

p_a: percentage of a^{th} airline share by total number of seats between origin (i) airport.

Market concentration or competition for each o-d airport pair can be measured by calculating the Herfindahl Index (HI) (3). Herfindahl Index (HI) is the sum of the square of the percentages of the available seats by each airline (pa) in a particular market segment.

$$HI = {p_1}^2 + {p_2}^2 + {p_3}^2 + \ldots + {p_a}^2 = \sum_a {p_a}^2 \tag{5.4}$$

The value of Herfindahl Index varies from 0 to 1.Figure 5.2 shows the value of Herfindahl Index for separate market segments. A value of 1 corresponds to a monopoly; 0.5 corresponds to an industry with two equal-sized firms, 0.33 corresponds to an industry with three equal sized firms and so on and so forth (3). As a rule, any market having a Herfindahl Index greater than 0.4 is considered a highly concentrated market and less than 0.18 a less concentrated market. The higher the concentration, the more likely the fare will increase in that market segment.

3) Passenger Flows:

One way to understand the supply and demand variable is using the passenger flows between the o-d pair on an aggregate level. A large number of passenger flows tend to reduce the average fare. However this may not be true for certain cases having higher Herfindahl Index values, thereby increasing the average fare. Hence causality between Herfindahl Index and passenger flow was observed.

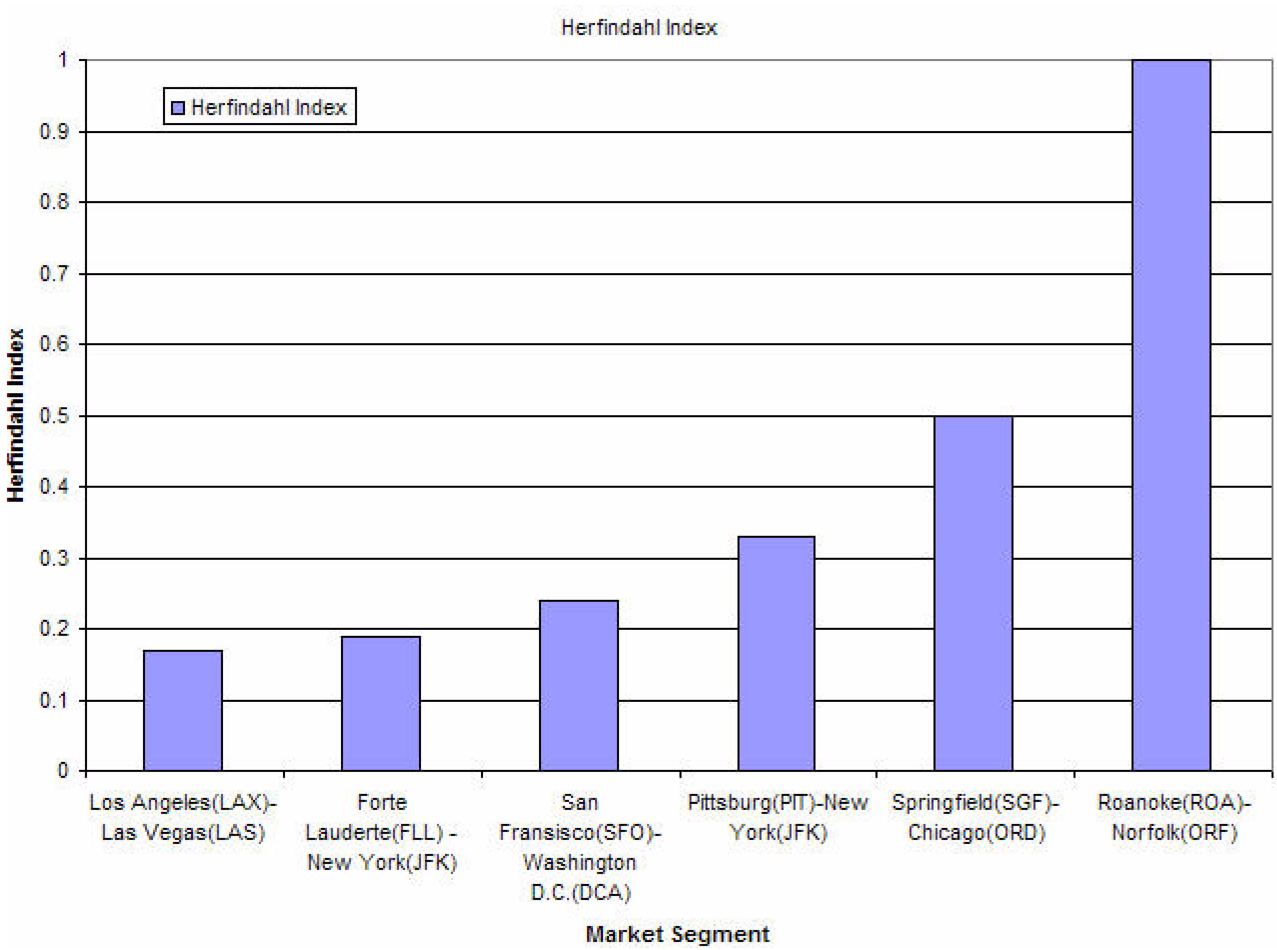


Figure 5.2 Herfindahl Index for Different Market Segments.

4) Low Cost carrier presence:

It is a general trend that presence of low cost carrier will tend to reduce the average fare between an o-d pair. Low Cost carriers usually don't offer business service and only offer point-to-point service, thereby reducing their operating costs. A list of six low cost carriers was identified from Bureau of Transportation Statistics (8). The fare per mile of all o-d airport pairs without the presence of low cost carriers and all o-d airport pairs with the presence of low cost carrier is plotted together as shown in the Figure 5.3. The figure clearly shows that o-d pairs with low cost carrier presence tend to pay less.

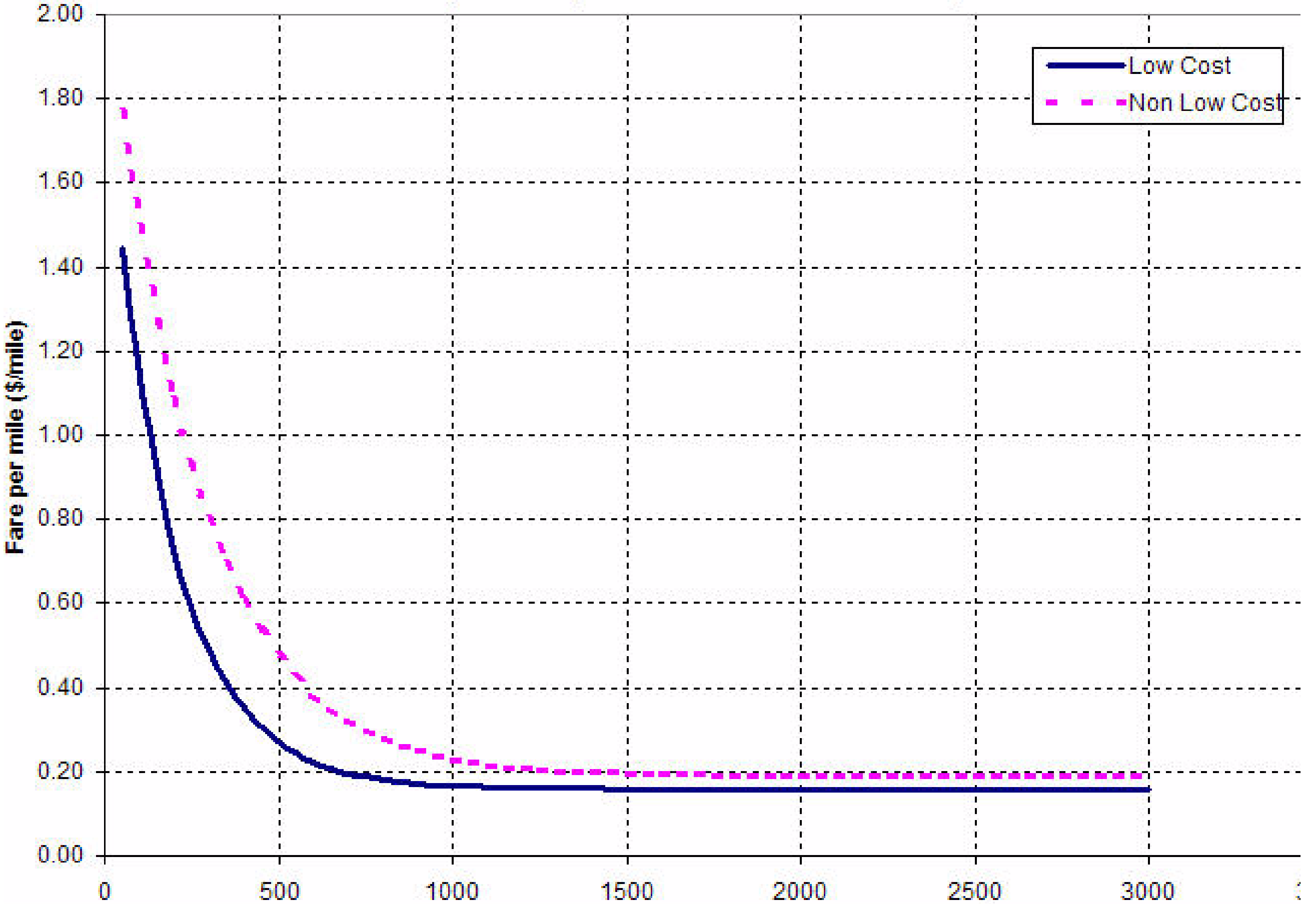


Figure 5.3 Fare per Mile for Low Cost vs. Legacy Carriers.

5) Origin and Destination Type:

In our analysis, airports are classified into the following types depending on the number of enplanements: Large Hub - 1, Medium Hub - 2, Small Hub - 3 and Non Hub - 4.

It is usually believed that traveling between Large Hubs is inexpensive than traveling from other airports. Also on a macro level the overall supply and demand, expenses and revenue would tend to drive the costs down in large airports.

5.2 Fare Models

Using all the causal variables mentioned above a family of '*Fare Model's*' was created for both Business and Coach Class. They are as follows:

1) Table Function:

The Table Function is the weighted average of the mean fare paid between 685 x 685 airports.

The mean fare for a single o-d pair is determined using the following formula

$$\mu = \sum(x \cdot p(x)) \tag{5.5}$$

Where,

x - Fare ($).

$p(x)$ - probability of x.

2) Non-Linear Regression Fare Model:

A generic fare per mile model is used to predict fare per mile only using the mean round trip distance traveled as an independent variable. It is a non-linear regression model also known as Harris Model. The model is given below:

$$y = (1/(a + b \cdot x^c)) \tag{5.6}$$

where, y is the fare per mile ($/mile), a, b and c are the model parameters, and x is the round trip distance in statue miles. This Harris Model is used in agricultural economics. Figure 5.4 shows the generic fare model for business class.

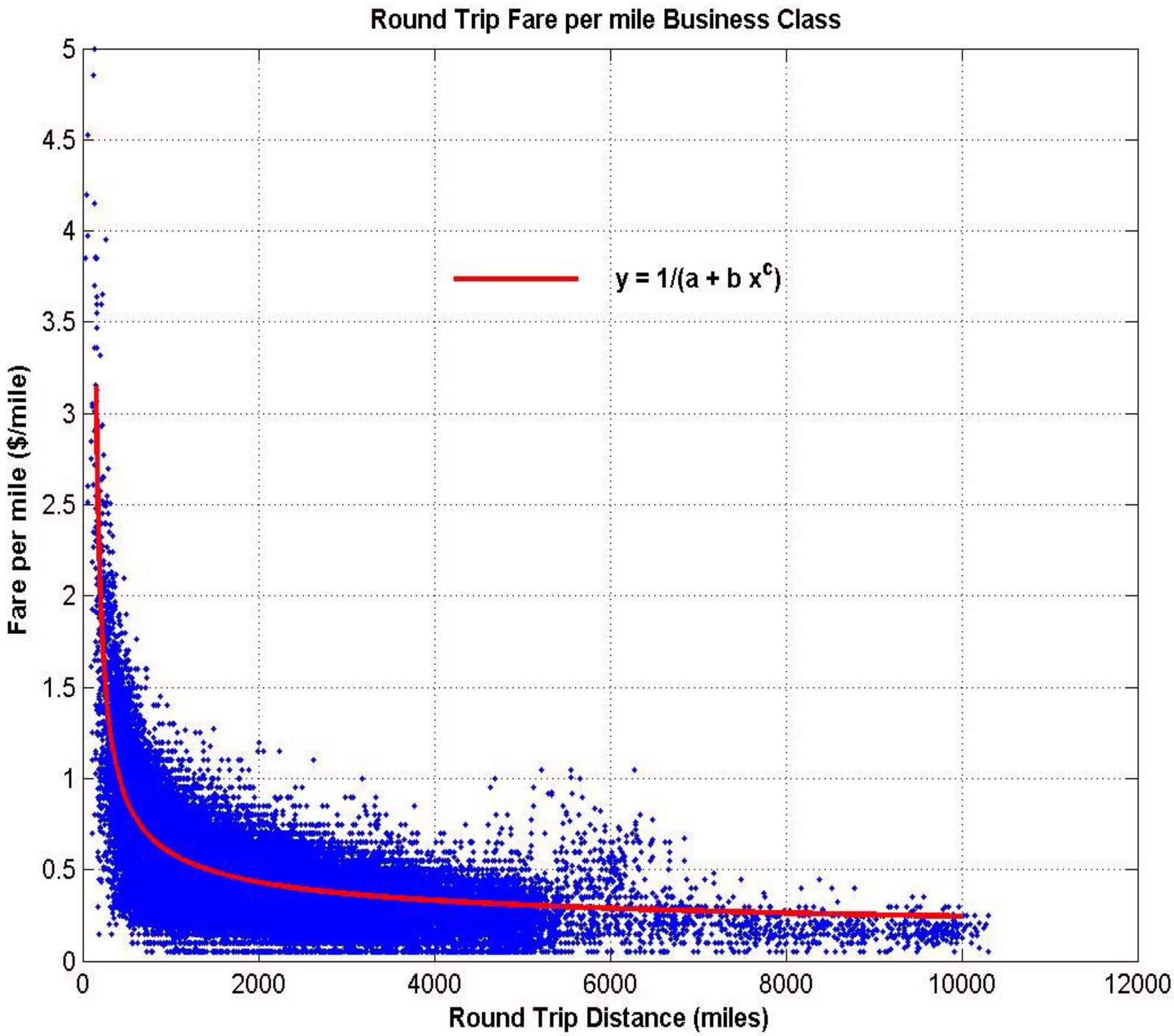


Figure 5.4 Non-linear Regression Fare Model.

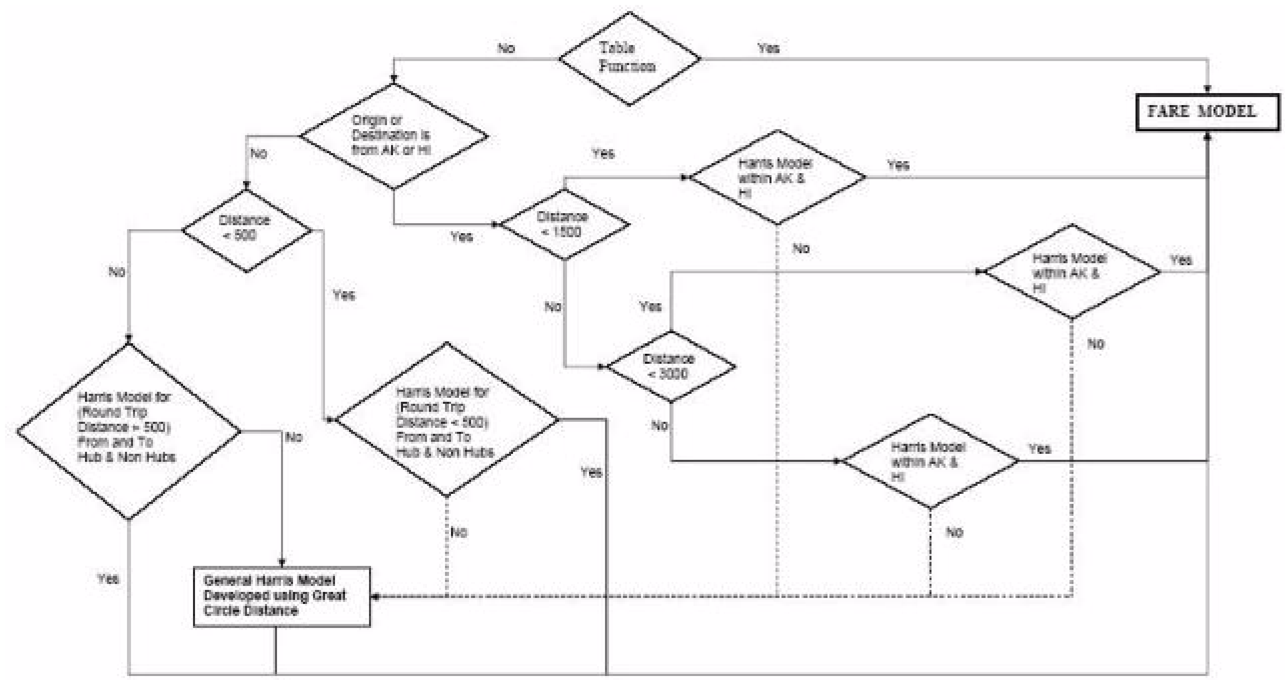


Figure 5.5 Fare Flow Model.

5.2.1 Fare Flow Model

A family of generic fare models was developed as an input for the *'Fare Flow Model'*.The *'Fare Flow Model'* is a combination of Table Function and the generic fare models. A flow chart was developed for this fare model. Figure 5.5 shows the Flow chart of the fare flow model.One of the categories of classifications within this model was trip length. The models were grouped together for different trip length distance categories. Description of the flow chart is given in the following steps:

- The model first checks for fare value between any o-d pair in Table function.
- If the fare value in Table Function is not available, checks whether one of the o-d airports is in Alaska or Hawaii.
- If Origin and Destination airport is in Alaska or Hawaii, Checks the distance.
- If distance is less than 1500 miles, then uses the '*Harris Model*' within AK & HI.

- If No, then checks if distance is greater than 1500 miles and less than 3000 miles and uses the '*Harris Model*' developed for that distance category.
- If the distance is greater than 3000 miles it uses the '*Harris Model*' for distance greater than 3000 miles.
- If the o-d airport is not Alaska or Hawaii it checks for o-d pair airports with distance less than 500 miles.
- If the distance is less than 500 miles, then uses the '*Harris Model*' developed for that category of distance and o-d pairs.
- If the distance is greater than 500 miles it uses the '*Harris Model*' for distance greater than 500 miles.
- Finally if o-d pair doesn't fall in any of the above category it uses the Generic Fare model developed using the Great Circle Distance.

Results from the model are shown in Table 5.3. These models have a predictability range from 30% all the way to 75%.

Table 5.3. Fare Flow Model Results.

Category	Fare Class	Distance	Number of o-d pairs	Model Parameters			R-square
				a	b	c	
Origin or Destionation pair in Alaska or Hawaii	Business	Distance < 1500	238	-3.06	2.47	0.09	0.6
	Business	Distance 1500-3000	138	0.0171	0.044	0.528	0.3
	Business	Distance 3000-10000	530	2.99	0.00015	1.072	0.3
	Coach	Distance <1500	295	-17.58	15.81	0.03	0.66
	Coach	Distance 1500-3000	250	-56.69	35.849	0.0688	0.41
	Coach	Distance 3000-10000	1940	-1.729	0.251	0.427	0.42
Origin or Destionation pair in Continental US	Business	Distance < 500	1236	-0.16259	0.084125	0.404494	0.467
	Business	Distance > 500	39439	-1.33164	0.242857	0.355303	0.688
	Coach	Distance < 500	1272	0.401088	0.001806	1.042827	0.565
	Coach	Distance > 500	70080	-3.86696	0.523035	0.371705	0.744
Fare models using Great Circle Distance	Business	Distance in GCD	40126	-3.21351	1.484201	0.176519	0.59
	Coach	Distance GCD	71329	-5.48925	1.156418	0.335299	0.73

5.2.2 Statistical Validation of '*Fare Flow Model*'

The Fare flow model was then tested using a non parametric statistical test for non-similarity between the generic fare models. The Wilcoxon rank-sum test is a nonparametric alternative to the two-sample t-test which is based solely on the order in which the observations from the two samples fall (14).

The results from Wilcoxon rank-sum hypothesis test can be summarized as follows:

- The result of the hypothesis test is *H*.
- *H* is 'zero' if the populations of two independent samples are not significantly different. *H* is 'one' if the two populations are significantly different.
- *p-value* is the probability of observing a result equally or more extreme than the one using the samples if the null hypothesis is true.
- If *p-value* is near 'zero', this casts doubt on this hypothesis.
- *Z* is the value of the normal (*Z*) statistic used to compute *p-value*.

Table 5.4. Wilcoxon Rank-Sum Test.

Category	Distance Group	Hypothesis	H-value	Z-value	p-value
Origin or Destionation pair in Alaska or Hawaii	1500	$\mu_{BUS} - \mu_{COACH} = 0$	1	5.5	3.00E-07
	1501-3000	$\mu_{BUS} - \mu_{COACH} = 0$	1	6.25	4.08E-10
	3001-10000	$\mu_{BUS} - \mu_{COACH} = 0$	1	6.64	3.02E-11
Origin or Destionation pair in Continental US	500	$\mu_{BUS} - \mu_{COACH} = 0$	1	6.34	2.21E-10
	501-ABOVE	$\mu_{BUS} - \mu_{COACH} = 0$	1	6.49	8.48E-11
Fare models using Great Circle Distance	200-Above	$\mu_{BUS} - \mu_{COACH} = 0$	1	5.7585	8.48E-09

The results from Wilcoxon rank-sumtest performed on the '*Fare flow models*' indicate that the models are dissimilar and are independent from each other. The *p-values* imply that the models are statistically significant.

5.2.3 Multiple Linear Regression Model

To test the hypotheses about the factors that affect the cost of air travel; multiple regression equations were undertaken on the basis of fare class.

Coach Class Fare Analysis:

The general regression model for coach class fares is given below:

$$f_{cij} = \beta_0 + \beta_1 d_{ij} + \beta_2 p_{cij} + \beta_3 h_i + \beta_4 h_j + \beta_5 lc_{ij} + \beta_6 o_i + \beta_7 d_j + e_{ij} \quad (5.7)$$

where,

f_{cij} : annual average round-trip fare for coach class between *i* and *j*.

d_{ij} : round trip distance in statue miles between *i* and *j*.

h_i : Herfindahl Index at the origin airport i.

h_j : Herfindahl Index at the origin airport j.

p_{cij}: annual coach class type passenger flows between *i* and *j*.

lc_{ij}: low cost carrier presence between *i* and *j* dummy variable 0 or 1.

o_i : origin airport type (*i*) [1, 2, 3, and 4].

d_j: destination airport type (*j*) [1, 2, 3, and 4].

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$: model parameters to be estimated.

e_{ij} - residual.

The results of ordinary least square (OLS) regression analysis,for coach class, are shown in Table 5.5.The parameter, average distance, has a positive sign, showing that longer trips have more average fare value. Competition is one of the main causes that affect airfares. The higher the competition, fares tends to be lower. The positive sign on the competition parameters, herfindahl index at the origin and destination airport, indicate that lesser the competition more the average fare between the o-d pair. It also shows that the competition at the destination airport is more critical than the competition at the origin airport.The annual passenger flows are higher between larger airport pairs. This flow is one of the main reason for congestion in these large airports; leading to more indirect operating costs. These costs are directly passed on to the passengers leading to higher fares, as indicated by the positive sign on annual average passenger flows.Low-cost carriers have completely changed the scenario of air travel in the US. These airlines have a successful business model to reduce indirect operating costs, thereby offering cheaper fares. Any presence of low-cost carrier at the origin airport tends to reduce the average fare. This is indicated by the negative sign of the causal variable low-cost carrier presence.The origin and destination airport type variables both have positive affects, suggesting that airfare tends to be higher at smaller airports. Again, the destination airport type is more critical that the origin airport type.

Table 5.5. Parameter Estimate for Coach Fare Class Regression, Average Coach Fare (fij).

	Parameter Estimate	t-statistic	p-value	Variance Inflation Factor
Intercept	212.978	76.027	0.000	
Annual Passenger Flow (pij)	0.00022	4.363	0.000	1.324
Average Distance (Dij)	0.055	150.750	0.000	1.039
Herfindahl Index at Origin Airport i	14.984	6.771	0.000	1.362
Herfindahl Index at Destination Airport j	25.089	11.320	0.000	1.360
Low Cost Carrier Presence at Origin Airport i	-63.310	-29.662	0.000	1.201
Origin Airport Type	3.825	6.443	0.000	1.612
Destination Airport Type	8.577	14.306	0.000	1.643
Adjusted R-square	0.516			
Number of Observations	23,599			

The p-values of the causal variables indicates that the variables are statistically significant. Variance Inflation Factor (VIF) is used as a measure of multicollinearity diagnostic. Multicollinearity implies near-linear dependence among the regressors (13).The VIF for each term in the model measures the combined effect of the dependences among the regressors on the variance of that term. The VIF values larger than ten imply serious problems with multicollinearity. VIFs for the regression model indicate that the model is not multicollinear and the regressors are independent of each other.

A scatter plot of standardized residuals on the standardized predicted values was plotted to detect outliers and non-linearity. The plot is shown in Figure 5.6. The plot fans out in a funnel shape, indicating a presence of heteroscedasticity. This issue was addressed by taking natural logarithm transformation on all but the dummy variables.

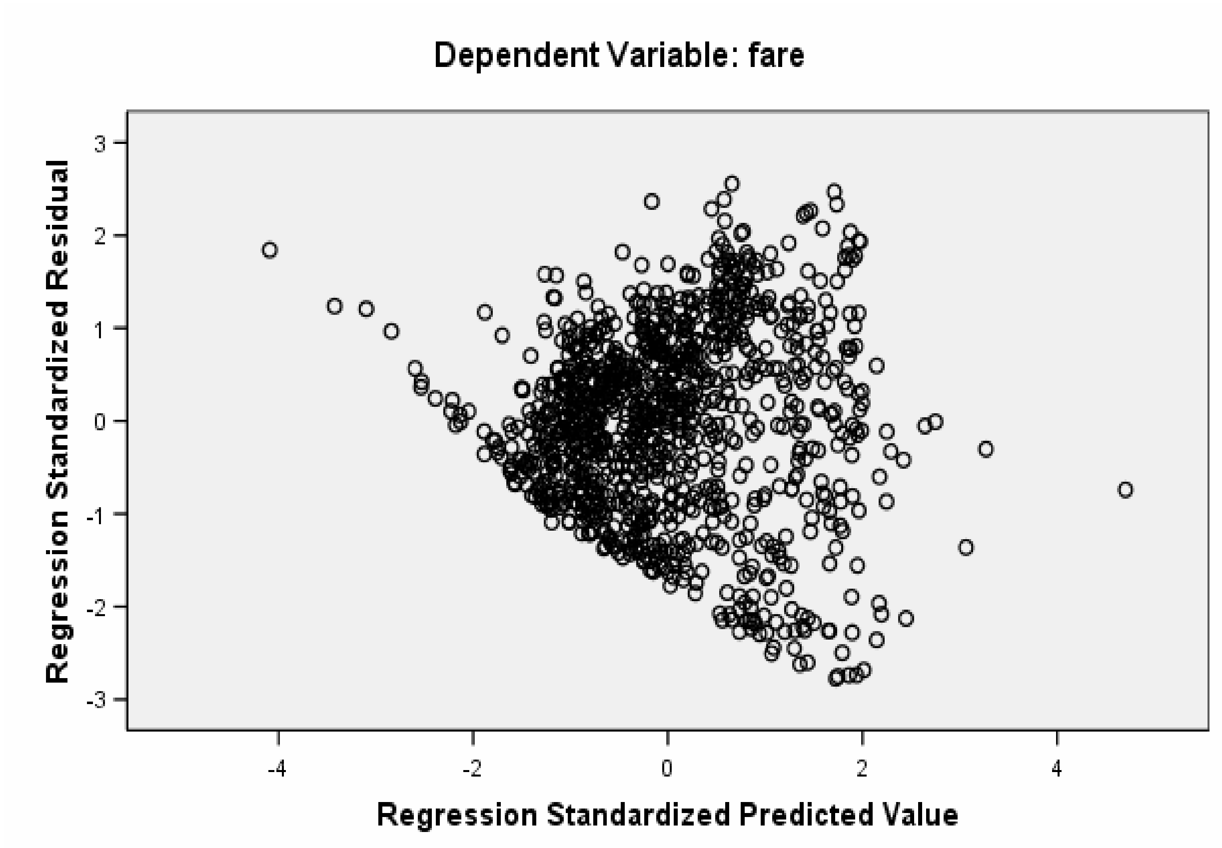


Figure 5.6 Scatter Plot of Standardized Residual and Predicted Values (fare).

The regression estimates and residual plot from the transformed logarithm model are shown in Table 5.6 and Figure 5.7. Table 5.6 is consistent with Table 5.5 for regressor significance.The p-values of the regressor variables indicates that the variables are statistically significant.VIFs for the regression model indicate that the model is not multicollinear and the regressors are independent of each other.

The residual plot of logarithm-in-linear regression does not show any clear pattern, indicating no presence of heteroscedasticity. This is further statistically verified by White's test for heteroscedasticity.

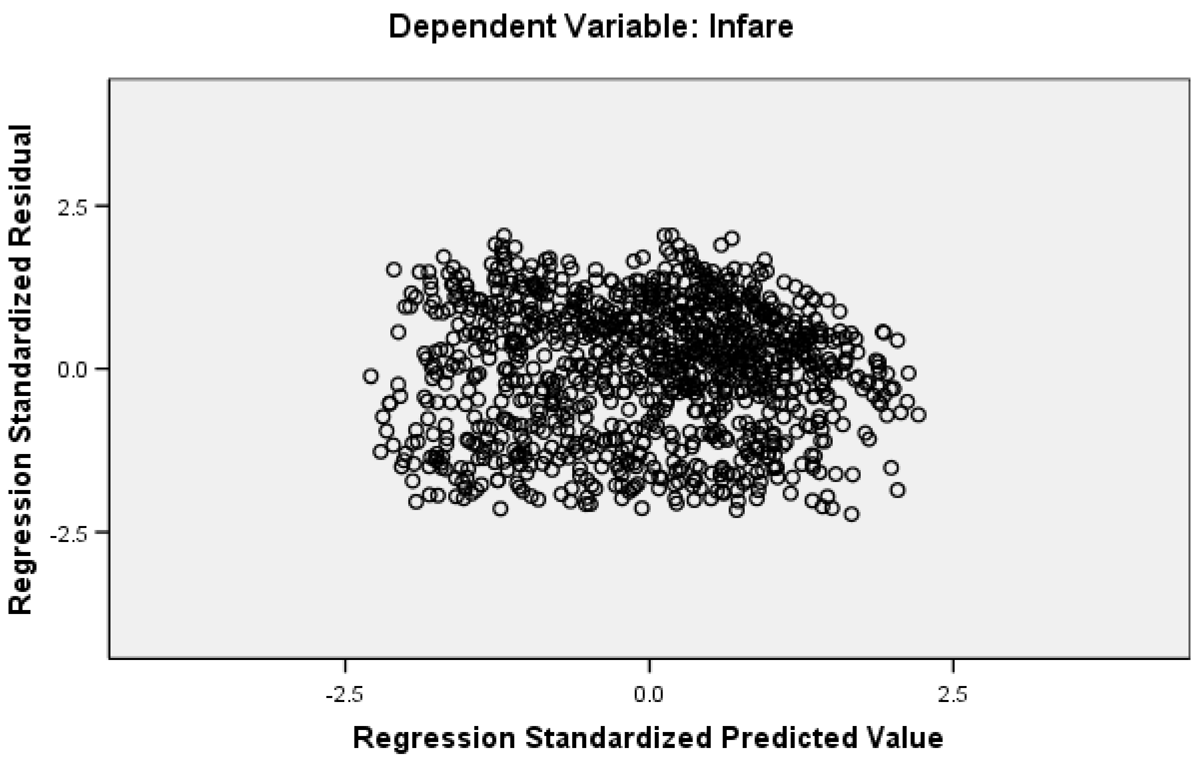


Figure 5.7 Scatter Plot of Standardized Residual and Predicted Values(ln(fare) - Coach Class).

Table 5.6. Parameter Estimate for Coach Fare Regression,Logarithm of Coach Fare (fij).

	Parameter Estimate	t-statistic	p-value	Variance Inflation Factor
Intercept	4.045	152.271	0.000	
Ln(Annual Passenger Flow (pij))	0.034	27.410	0.000	1.785
Ln(Average Distance (Dij))	0.297	126.648	0.000	1.081
Ln(Herfindahl Index at Origin Airport i)	0.007	3.701	0.000	1.468
Ln(Herfindahl Index at Destination Airport j)	0.016	8.217	0.000	1.433
Ln(Low Cost Carrier Presence at Origin Airport i)	-0.172	-32.460	0.000	1.178
Ln(Origin Airport Type)	0.028	15.352	0.000	1.416
Ln(Destination Airport Type)	0.019	10.139	0.000	1.530
Adjusted R-square	0.483			
Number of Observations	23,599			

Table 5.7. White's test for Heteroscedasticity for Coach Fare Regression.

	White's test		
	Standard Errors	t-statistic	p-value
Intercept	0.0700	150.27	0.000
Ln(Annual Passenger Flow (pij))	0.0080	27.41	0.000
Ln(Average Distance (Dij))	0.0030	126.65	0.000
Ln(Herfindahl Index at Origin Airport i)	0.0050	3.70	0.000
Ln(Herfindahl Index at Destination Airport j)	0.0050	8.22	0.000
Ln(Low Cost Carrier Presence at Origin Airport i)	0.0200	-32.46	0.000
Ln(Origin Airport Type)	0.0040	15.35	0.001
Ln(Destination Airport Type)	0.0010	10.14	0.000

White (6) developed an algorithm for correcting the standard errors in OLS when heteroscedasticity is present. The results from White's test is shown in Table 5.7. The t-statistic for White's test is identical to the t-values from the logarithm-in-linear OLS regressor variables.The greater the heteroscedasticity, the larger the difference between the OLS t-values and the White test t-values. Since, the t-values are identical, the model is homoscedastic.Appendix A has the SPSS syntax used to develop the White's heteroscedasticity test results.

<u>Business Class Fare Analysis:</u>

The general regression model for this approach is given below:

$$f_{bij} = \beta_0 + \beta_1 d_{ij} + \beta_2 p_{bij} + \beta_3 h_i + \beta_4 h_j + \beta_5 o_i + \beta_6 d_j + e_{ij} \quad (5.8)$$

where,

f_{bij} : annual average round-trip fare for business class between i and j.

d_{ij} : round trip distance in statue miles between i and j.

h_i : Herfindahl Index at the origin airport i.

h_j : Herfindahl Index at the origin airport j.

p_{bij}: annual business class type passenger flows between i and j.

o_i : origin airport type (i) [1, 2, 3, and 4].

d_j: destination airport type (j) [1, 2, 3, and 4].

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$: model parameters to be estimated.

e_{ij} - residual.

The results of ordinary least square regression analysis,for business class, are shown in Table 5.8.The parameter, average distance, has a positive sign, showing that longer business trips have more average fare value. The positive sign on the competition parameters, herfindahl index at the origin and destination airport, indicate that competition also affects business class fares; lesser the competition more the average fare between the o-d pair. It also shows that the competition at the origin airport is more critical than the competition at the destination airport in case of business class trips. The annual passenger flows variable has a contradictory affect on the business class fares; indicating that an increase in business class passengers flows tend to reduce the overall operating costs thereby reducing the average fare. The origin and destination type causal variables have a negative affect on the business fares, implying that the business fares tends to be lower at small sized airports.

The p-values of the causal variables indicates that the variables are statistically significant. VIFs for the regression model indicate that the model is not multicollinear and the regressors are independent of each other.

Table 5.8. Parameter Estimate for Business Fare Class Regression, Average Business Fare.

	Parameter Estimate	t-statistic	p-value	Variance Inflation Factor
Intercept	546.562	29.782	0.000	
Annual Passenger Flow (pij)	-0.01897	-22.444	0.000	1.235
Average Distance (Dij)	0.145	50.061	0.000	1.124
Herfindahl Index at Origin Airport i	181.501	12.190	0.000	1.639
Herfindahl Index at Destination Airport j	179.829	12.204	0.000	1.620
Origin Airport Type	-51.464	-9.994	0.000	1.732
Destination Airport Type	-34.361	-6.639	0.000	1.759
Adjusted R-square	0.361			
Number of Observations	6,517			

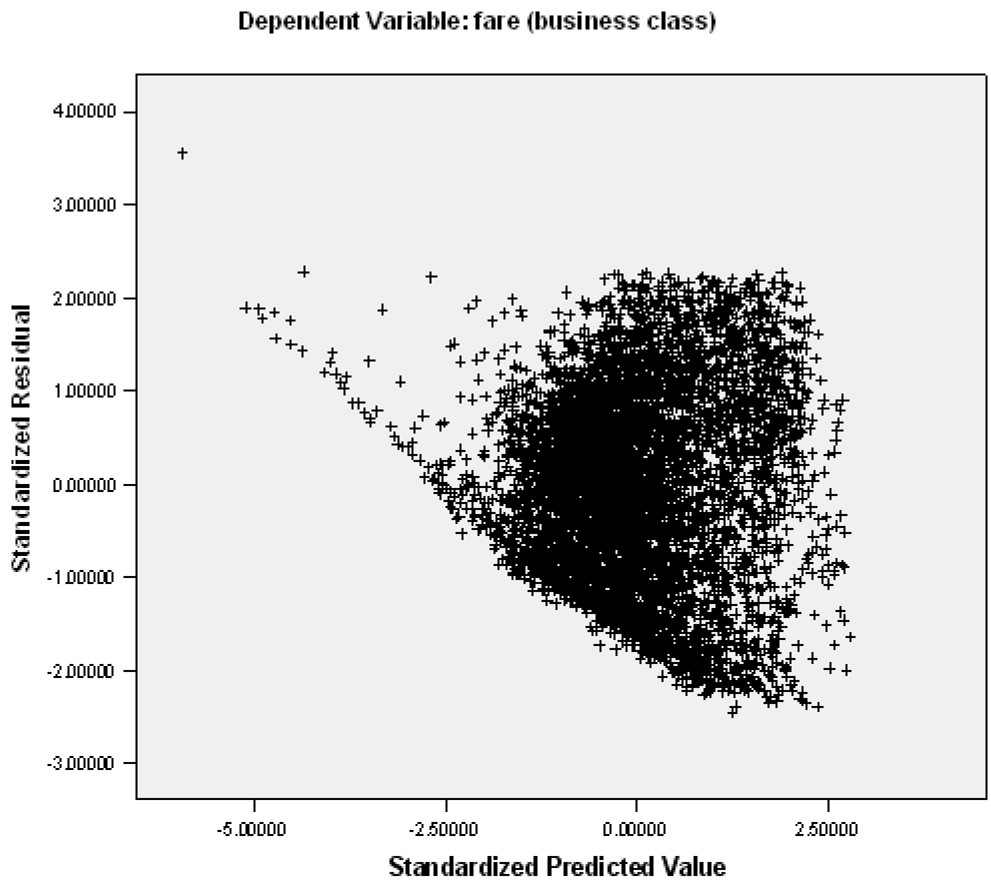


Figure 5.8 Scatter Plot of Standardized Residual and Predicted Values (fare - Business Class).

A scatter plot of standardized residuals on the standardized predicted values was plotted to detect outliers and non-linearity. The plot is shown in Figure 5.8. The plot fans out in a funnel shape, indicating a presence of heteroscedasticity. This issue is addressed by taking natural logarithm transformation on all but the dummy variables.

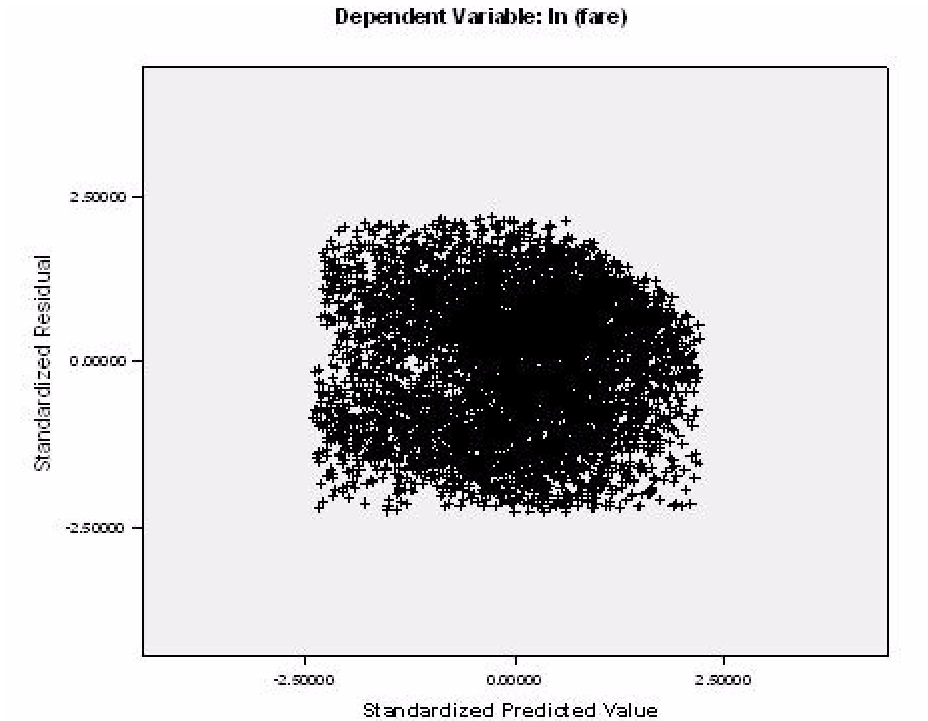


Figure 5.9 Scatter Plot of Standardized Residual and Predicted Values(ln(fare) - Business Class).

The regression estimates and residual plot from the transformed logarithm model are shown in Table 5.9 and Figure 5.9. Table 5.9 is consistent with Table 5.8 for regressor significance.The p-values of the regressor variables indicates that the variables are statistically significant.VIFs for the regression model indicate that the model is not multicollinear and the regressors are independent of each other.

The residual plot of logarithm-in-linear regression does not show any clear pattern, indicating no presence of heteroscedasticity. This is further statistically verified by White's test for heteroscedasticity.

Table 5.9. Parameter Estimate for Business Fare Class Regression, Logarithm of Business Fare (fij).

	Parameter Estimate	t-statistic	p-value	Variance Inflation Factor
Intercept	6.450	81.415	0.000	
Ln(Annual Passenger Flow (pij))	-0.178	-47.589	0.000	1.871
Ln(Average Distance (Dij))	0.271	35.177	0.000	1.160
Ln(Herfindahl Index at Origin Airport i)	0.054	8.945	0.000	1.506
Ln(Herfindahl Index at Destination Airport j)	0.047	7.883	0.000	1.490
Ln(Origin Airport Type)	-0.173	-25.989	0.000	1.953
Ln(Destination Airport Type)	-0.146	-22.365	0.000	1.904
Adjusted R-square	0.427			
Number of Observations	6,517			

Table 5.10. White's test for Heteroscedasticity for Business Fare Regression.

	White's test		
	Standard Errors	t-statistic	p-value
Intercept	0.1000	81.415	0.000
Ln(Annual Passenger Flow (pij))	0.0100	-47.589	0.000
Ln(Average Distance (Dij))	0.0060	35.177	0.000
Ln(Herfindahl Index at Origin Airport i)	0.0090	8.945	0.000
Ln(Herfindahl Index at Destination Airport j)	0.0090	7.883	0.000
Ln(Origin Airport Type)	0.0090	-25.989	0.000
Ln(Destination Airport Type)	0.0090	-22.365	0.000

The results from White's test is shown in Table 5.10. The t-statistic for White's test is identical to the t-values from the logarithm-in-linear OLS regressor variables.The greater the heteroscedasticity, the larger the difference between the OLS t-values and the White test t-values. Since, the t-values are identical, the OLS regression model for business class fares is homoscedastic.

CHAPTER 6 Model Validation & Model Implementation

Validation statistics measure the error or accuracy of model prediction.In order to test the models ability to forecast, validation requires comparison of the model predictions with a randomized sample of the original data set. In the previous section of this thesis, the fare models were successfully tested for statistical inference and significance.In this penultimate section of the thesis, the fare models are finally validated using standard statistical validation techniques. After validation testing, the model is successfully implemented in the Transportation Systems Analysis Model (TSAM) for SATS.

6.1 Model Validation

Validation statistics measure the error or accuracy of model prediction.In order to test the models ability to forecast, validation requires comparison of the model predictions with a randomized sample of the original data set. A forty percent cross-section sample from the actual data set, for quarter one of the year 2001, is compared with the model predictions.

The mean-squared error is one of the most commonly used measures of success for numeric prediction. This value is computed by taking the average of the squared differences between each computed value and its corresponding correct value. The root mean-squared error is simply the square root of the mean-squared-error. The root mean-squared error gives the error value the same dimensionality as the actual and predicted values. The RMSE measures the deviation between the model predicted air fares and the actual itinerary fare from the database record and is given as:

$$Percent(RMSE) = \left(\left(\sqrt{\left(\sum_j (Predictedfare_j - Actualfare_j)\right)^2 / (NOB-1)}\right) / \left(\left(\sum_j Actualfare_j\right) / (NOB)\right)\right) \times 100 \tag{6.1}$$

where, *j - jth observation number.*

NOB - Number of Observations.

A large percent RMSE indicates a large deviation between the forecasted model air fare and the database fare; whereas, a small percent RMSE indicates a small deviation between the predicted air fare and the actual fare. The percent RMSE for Coach and Business Class fares for origin and destination airport type by different distance groups is shown in Table 6.1& Table 6.2. The RMSE value for coach fare estimation process varies from 40% to 53%. The model prediction capability for coach class fares from non-hub to medium-hub is very weak. This aspect of model prediction ability must be improved in later studies. The RMSE value for business fare estimation process varies from 40% to 63%. Since, no previous studies have been performed validating a generic cost model, these RMSE values could be used as a benchmark to improve the accuracy of model prediction.

Table 6.1. RMSE results for coach class fares.

Distance Category	O\D	Number of Observations: Hub	Medium-Hub	Small-Hub	Non-hub
100-1500	Hub	17334	8314	7589	4181
	Medium-Hub	8487	3371	3361	1890
	Small-Hub	8984	3686	2931	1298
	Non-hub	4933	2144	1327	343
	O\D	% RMSE: Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	54	54	52	47
	Medium-Hub	53	50	49	47
	Small-Hub	51	48	49	47
	Non-hub	47	46	48	44

Distance Category	O\D	Number of Observations: Hub	Medium-Hub	Small-Hub	Non-hub
1501-3000	Hub	33959	14910	7208	5048
	Medium-Hub	16844	7022	3486	2163
	Small-Hub	12886	4894	2344	1039
	Non-hub	7101	2550	841	186
	O\D	% RMSE: Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	52	48	47	42
	Medium-Hub	48	41	44	41
	Small-Hub	42	41	44	41
	Non-hub	41	40	40	35

Distance Category	O\D	Number of Observations: Hub	Medium-Hub	Small-Hub	Non-hub
3001-5000	Hub	30055	9357	3180	2768
	Medium-Hub	11787	3771	1360	674
	Small-Hub	6063	1885	431	204
	Non-hub	3155	793	141	34
	O\D	% RMSE: Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	53	52	45	39
	Medium-Hub	49	48	45	40
	Small-Hub	41	42	40	38
	Non-hub	40	41	43	36

Distance Category	O\D	Number of Observations: Hub	Medium-Hub	Small-Hub	Non-hub
ALL	Hub	81348	32581	17977	11997
	Medium-Hub	37118	14164	8207	4727
	Small-Hub	27933	10465	5706	2541
	Non-hub	15189	5487	2309	563
	O\D	% RMSE: Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	53	51	49	43
	Medium-Hub	50	45	46	43
	Small-Hub	45	44	47	43
	Non-hub	42	42	45	40

Table 6.2. RMSE results for business class fares.

Distance Category	O\D	Number of Observations			
		Hub	Medium-Hub	Small-Hub	Non-hub
100-1500	Hub	14269	8092	6117	2299
	Medium-Hub	8082	6046	3344	506
	Small-Hub	6058	3747	2035	380
	Non-hub	2186	555	414	49
		% RMSE			
	O\D	Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	50	54	49	46
	Medium-Hub	55	57	54	48
	Small-Hub	51	55	51	43
	Non-hub	48	44	45	42

Distance Category	O\D	Number of Observations			
		Hub	Medium-Hub	Small-Hub	Non-hub
1501-3000	Hub	12303	7130	3406	690
	Medium-Hub	9169	6356	1927	163
	Small-Hub	5363	2449	589	40
	Non-hub	919	217	38	0
		% RMSE			
	O\D	Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	53	57	54	48
	Medium-Hub	55	65	59	52
	Small-Hub	55	56	51	64
	Non-hub	52	53	49	0

Distance Category	O\D	Number of Observations			
		Hub	Medium-Hub	Small-Hub	Non-hub
3001-5000	Hub	9287	3644	1075	235
	Medium-Hub	4846	3138	720	33
	Small-Hub	1753	852	65	3
	Non-hub	332	55	2	1
		% RMSE			
	O\D	Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	38	52	47	33
	Medium-Hub	54	62	59	44
	Small-Hub	51	54	51	43
	Non-hub	41	37	55	10

Distance Category	O\D	Number of Observations			
		Hub	Medium-Hub	Small-Hub	Non-hub
ALL	Hub	35859	18866	10598	3224
	Medium-Hub	22097	15540	5991	702
	Small-Hub	13174	7048	2689	423
	Non-hub	3437	827	454	50
		% RMSE			
	O\D	Hub	Medium-Hub	Small-Hub	Non-hub
	Hub	47	56	51	45
	Medium-Hub	56	63	58	49
	Small-Hub	53	56	51	46
	Non-hub	49	46	46	40

A result from Flow Model is shown in Figure 6.1. It captures the variation in fare paid and distance travelled for coach fare from San Francisco (SFO) to Atlanta (ATL). The model predicts a mean fare of 639 $. The mean fare from the database observations was 620 $for the same origin- destination pair for the same fare category.

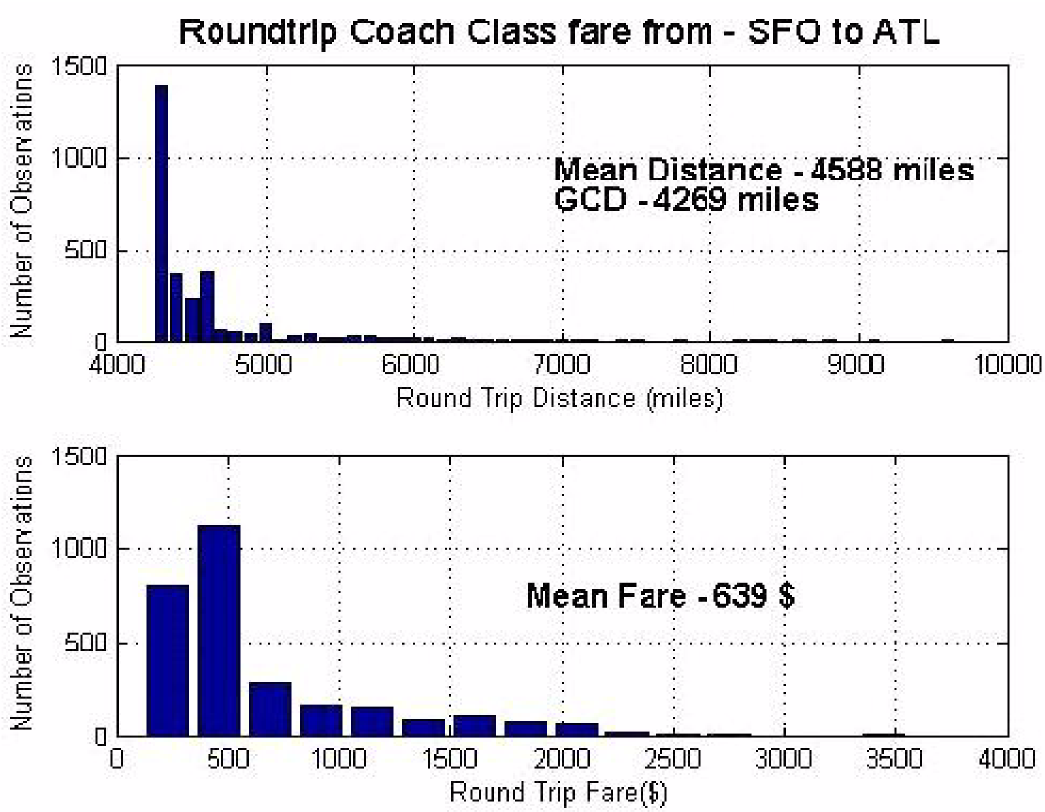


Figure 6.1 Sample result from Fare flow model.

6.2 Model Implementation

The validated and tested models developed in this thesis are implemented in the TSAM for

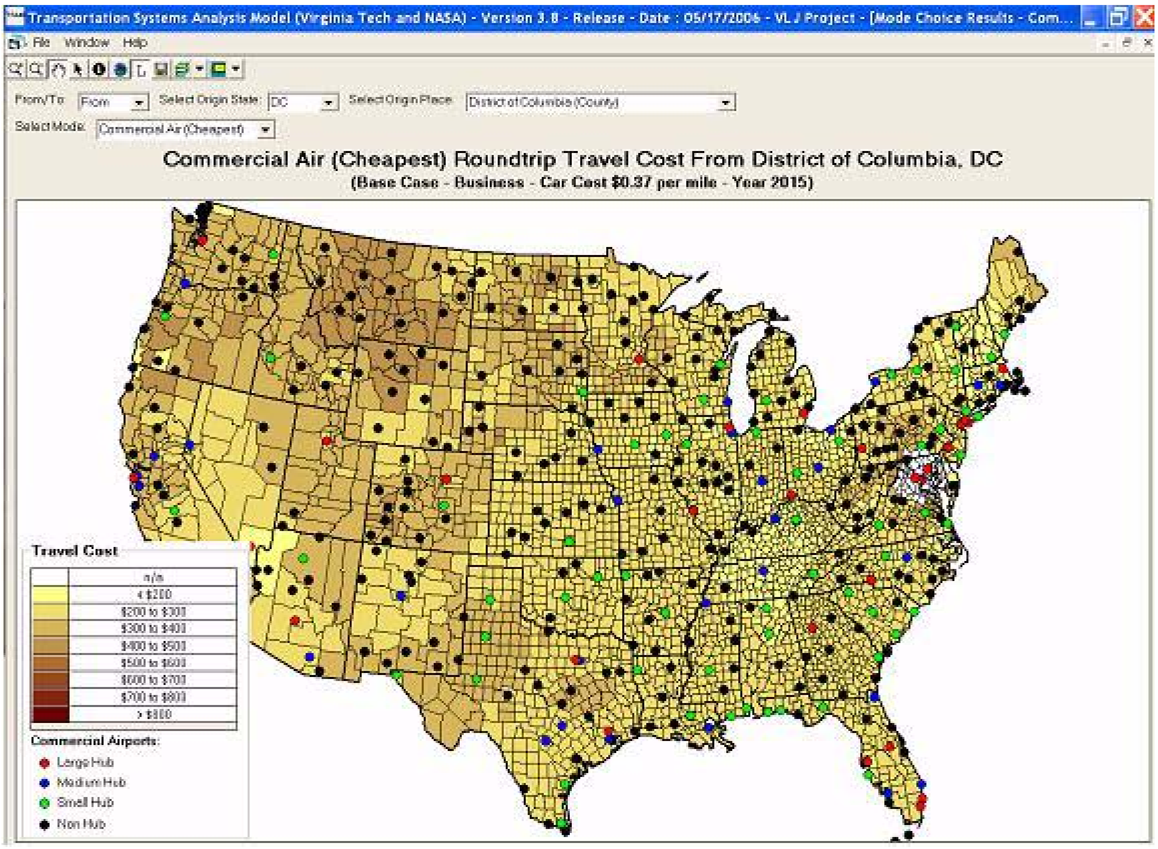
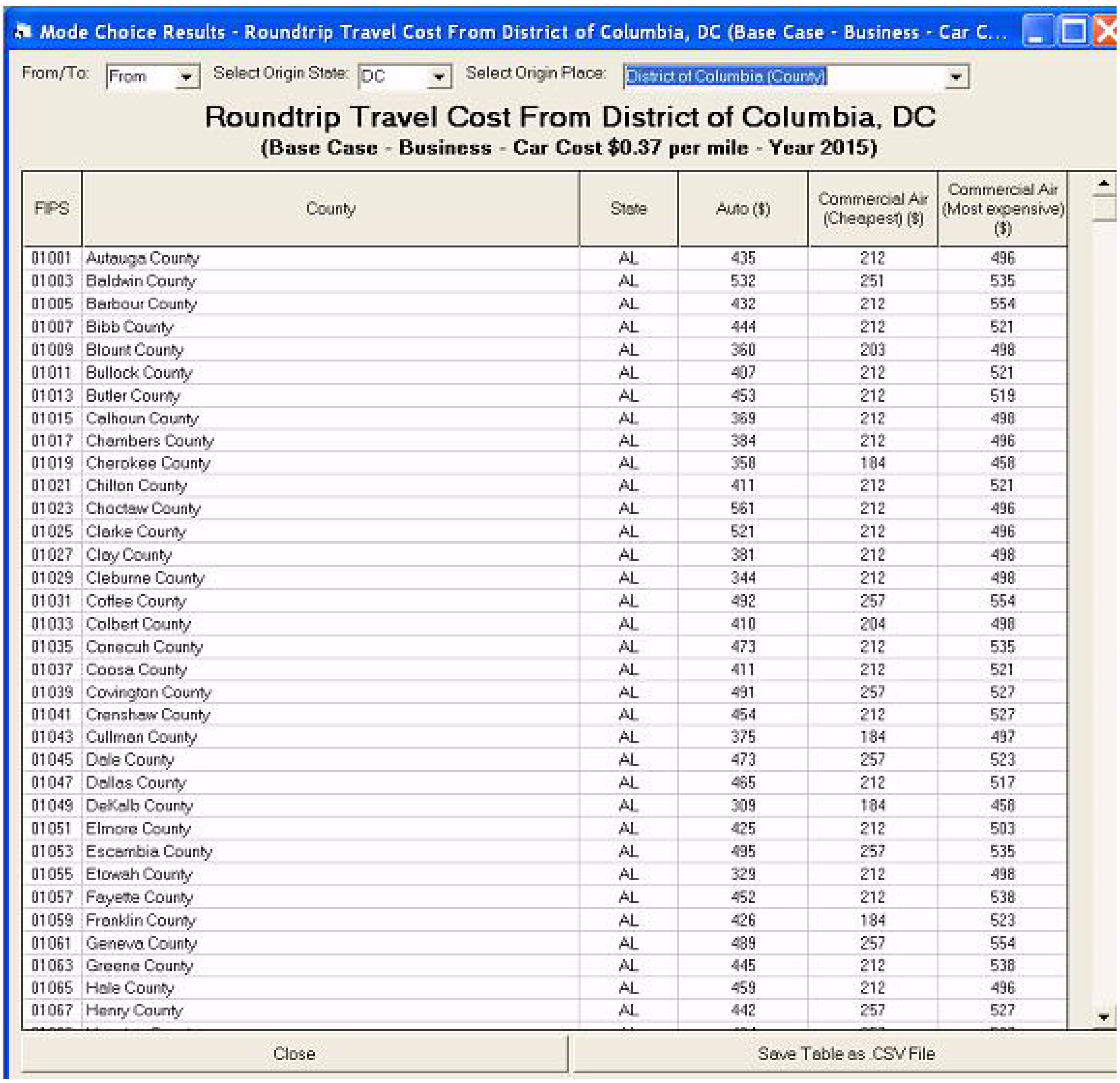


Figure 6.2 Commercial Air Travel Cost (Cheapest Fare). Focal County is D.C. Airports are Shown to Illustrate the Various Layers Available in TSAM.

SATS. This model is used as a measure of travel cost, a critical input, in the mode choice calibration of commercial air travel; a component in air travel alternative in SATS.TSAM software GUI has the ability to generate maps on a national scale for different model outputs such as commercial cost of air travel. Figure 6.2 is one such illustration. Figure 6.2 shows the cheapest cost of air travel from Washington D.C. to all regions in continental US.

TSAM also has the ability to generate tables based on different queries. Figure 6.3 shows a query of round trip travel cost, both coach and business class, from Washington D.C to all other counties in the US.



FIPS	County	State	Auto ($)	Commercial Air (Cheapest) ($)	Commercial Air (Most expensive) ($)
01001	Autauga County	AL	435	212	496
01003	Baldwin County	AL	532	251	535
01005	Barbour County	AL	432	212	554
01007	Bibb County	AL	444	212	521
01009	Blount County	AL	360	203	498
01011	Bullock County	AL	407	212	521
01013	Butler County	AL	453	212	519
01015	Calhoun County	AL	369	212	498
01017	Chambers County	AL	384	212	496
01019	Cherokee County	AL	358	184	458
01021	Chilton County	AL	411	212	521
01023	Choctaw County	AL	561	212	496
01025	Clarke County	AL	521	212	496
01027	Clay County	AL	381	212	498
01029	Cleburne County	AL	344	212	498
01031	Coffee County	AL	492	257	554
01033	Colbert County	AL	410	204	498
01035	Conecuh County	AL	473	212	535
01037	Coosa County	AL	411	212	521
01039	Covington County	AL	491	257	527
01041	Crenshaw County	AL	454	212	527
01043	Cullman County	AL	375	184	497
01045	Dale County	AL	473	257	523
01047	Dallas County	AL	465	212	517
01049	DeKalb County	AL	309	184	458
01051	Elmore County	AL	425	212	503
01053	Escambia County	AL	495	257	535
01055	Etowah County	AL	329	212	498
01057	Fayette County	AL	452	212	538
01059	Franklin County	AL	426	184	523
01061	Geneva County	AL	489	257	554
01063	Greene County	AL	445	212	538
01065	Hale County	AL	459	212	496
01067	Henry County	AL	442	257	527

Figure 6.3 Query of Table with Commercial Air Fares (from DC).

TSAM was used to compare cost of different modes and develop maps for travel cost savings between different modes. Figure 6.4 is an illustration of travel cost saving between commercial air and auto from Washington D.C. to all other counties in continental US.

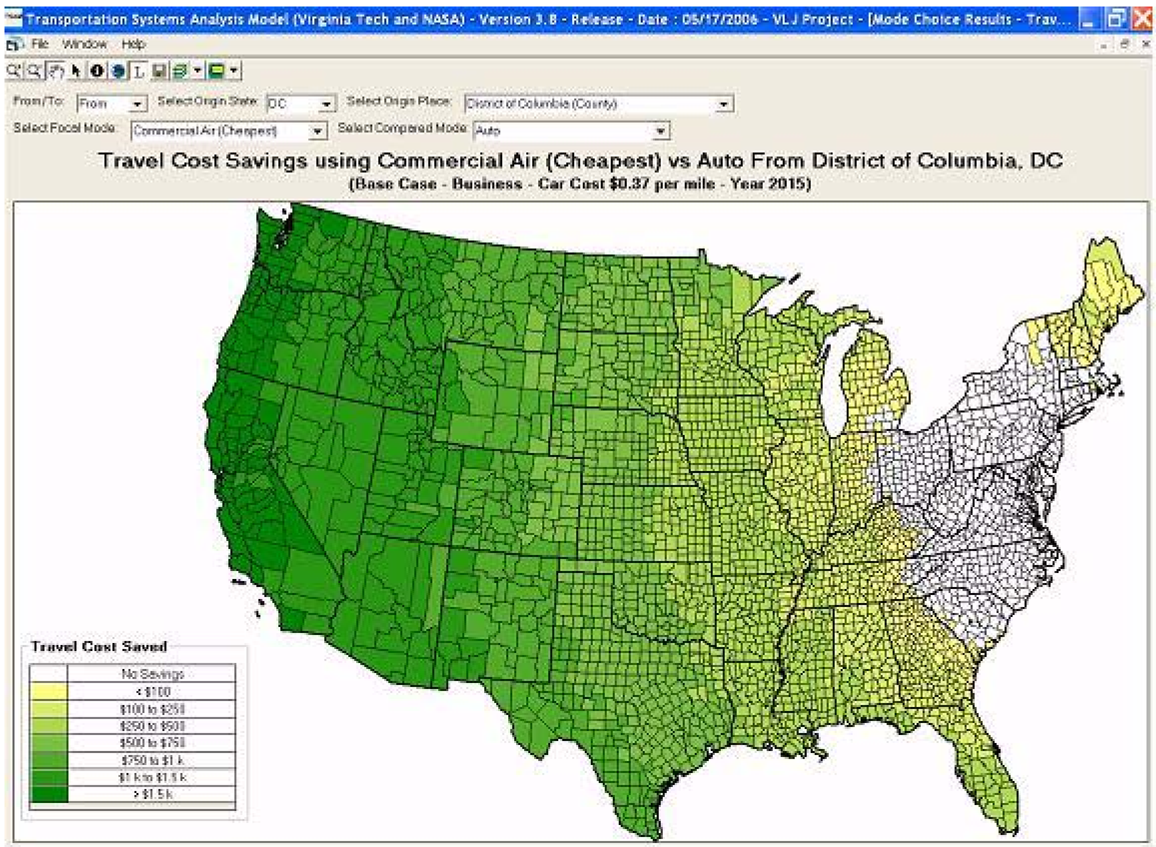


Figure 6.4 Travel Cost Savings (Airline vs. Auto Cost). Focal County is D.C.

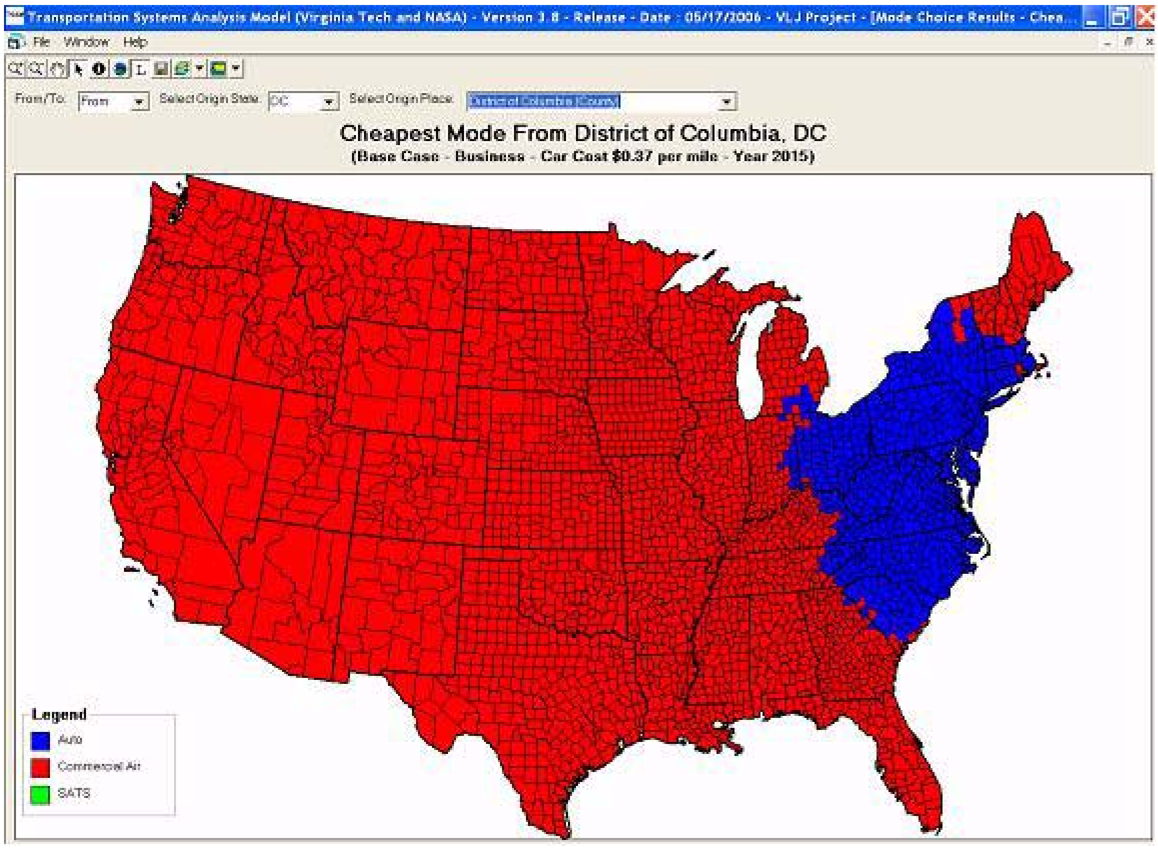


Figure 6.5 Travel Cost of Cheapest Mode Starting From Washington D.C.

Figure 6.5 shows the travel cost from Washington D.C. to other counties in the US by auto, commercial air and SATS.

In all of the above figures the 'Fare Models' developed in this thesis was used in computing the cost of commercial air travel.

CHAPTER 7 Conclusions & Recommendations

The central motivation of the thesis was to determine a generalized demand side cost metric input that can be applied on a national level for the mode choice analysis of SATS. For this purpose, the thesis developed a series of fare models, linear and non-linear, for different fare categories based on pre determined causal variables: distance, passenger flows, competition, low cost carriers in the market segment, and origin and destination type of the airports. These models were successfully tested for statistical inference and significance.The models were the validated using standard statistical validation techniques. After validation, the models were successfully implemented in Transportation Systems Analysis Model for SATS.

In this final section,the thesis concludes with inferences and recommendations.

7.1 Conclusions and Recommendations

7.1.1 Conclusion

The findings discussed in this thesis can be summarized as follows:

1)Long distance coach and business trips have higher average fare value. However, after deregulation, distance alone isn't the main fare factor determinant.

2)Competition is one of the main causes that affect airfares. The higher the competition, fares tends to be lower. Competition at the destination airport is more critical than the competition at the origin airport for coach class fares and vice a versa for business class fares.

3) The annual passenger flows are higher between large airport pairs. This flow is one of the main reason for congestion in these large airports; leading to more indirect operating costs. These costs are directly passed on to the passengers leading to higher coach class fares. The annual business class passenger flows has a contradictory affect on the business class fares; indicating that an increase in business class passengers flows tend to reduce the overall operating costs thereby reducing the average fare.

4) Low-cost carriers have completely changed the scenario of air travel in the US. These airlines have a successful business model to reduce indirect operating costs, thereby offering cheaper fares. This thesis introduces the presence of low cost carriers as a variable which affects air fares. Any presence of low-cost carrier at the origin airport tends to reduce the average fare.

5)The origin and destination airport type both have positive affects on coach class fares, suggesting that coach fares tends to be higher at smaller airports. Again, the destination airport type is more critical that the origin airport type for coach class fares. The origin and destination type causal variables have a negative affect on the business fares, implying that the business fares tends to be lower at small sized airports.

7.1.2 Recommendations

The results present in this thesis direct to a number of other vital possibilities for further research undertakings. They are:

1)Improving the model prediction abilities particularly in the non-hub to medium-hub market segment sector.

2)This model is an annual average estimate of the cost of nationwide air travel in the US. Further attempts of research can be made to address the seasonal variations in the cost of air travel.

3) Other attempts should be made to develop generic models i.e. harris models to estimate the cost of air travel between airport type to airport type.

4)Transportation demand estimation of SATS forecasts future traffic patterns and congestions in National Airspace System. This estimation process requires a time series based cost model metric to forecast future cost of air travel. A time series based cost model would be the next step in the same lines of this research.

With the events of September 11, 2001 and the attendant additional security and travel time required, reapplication and perhaps expansion of this modeling methodology is needed to show that this method herein described is needed to predict air fares after September 11,2001.

Appendix A

Department of Transportation Sample Data and Source Code

A.1 Commercial Service Database Analysis

A.1.1 Airline Origin and Destination Survey

a) Overview:

The Airline Origin and Destination Survey (DB1B) is a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. This database is used to determine air traffic patterns, air carrier market shares, and passenger, freight, and mail cargo flow.

b) Coverage :

The survey covers a 10% sample of airline tickets from reporting carriers.

c) Availability

First Year : 1994

Lat Year : 2002

Frequency : Quarterly

d) Data Tables

1) DB1BCoupon

2) DB1BMarket

3) DB1B Ticket

1) DB1BCoupon:

This table provides coupon-specific information for each itinerary, such as the operating carrier, origin and destination airports, number of passengers, fare class, coupon type, trip break indicator, gateway indicator, and distance.Figure A.1 is snapshot of the data available in Coupon table.

2) DB1BMarket :

This file contains (directional) origin and destination markets from the Origin and Destination Survey (DB1B), which is a 10% sample of airline tickets from reporting carriers. It includes such items as passengers, fares, and distances for each directional market, as well as information about whether the market was domestic or international. The file also reports operating and ticketing carrier information for flight segments within the directional market. This file is related to both the O&D Segment and Ticket files by the unique Market ID on each record.Figure A.2 shows the Market Data.

3)DB1BTicket:

This table contains summary characteristics of each domestic itinerary of the Origin and Destination Survey,including the reporting carrier,prorated market fare,number of market coupons,market miles flown,and carrier change indicators.Since Ticket data gave roundtrip distance and fare ,for the analysis Coupon ,Market and Ticket data tables were combined.Figure A.3 is a snapshot of the Ticket database.

4) Combined Coupon and Market:

We combine Coupon and Market database to create the main input data source for our analysis of oneway trips.The combined file is then saved as a ".dat" files which is then read in Matlab.Figure A.4 shows a sample of the combined database.

5) Combined Coupon, Market and Ticket:

The databases Coupon,Market and Ticket were combined to create the second main input data source for our analysis of round trips.This combined file is then saved as a ".dat" file which is then read in Matlab.Figure A.5 shows a sample of the combined database.

	break	dest	des	distance	farecl	itingeot	itinid	mktid	opcar	origin	origin	passenge	rpc
1	X	ATL	GA	1269	X	2	200016	200018	AA	ABQ	NM	1.00	AA
2	X	ABQ	NM	1269	X	2	200016	200019	AA	ATL	GA	1.00	AA
3	X	ATL	GA	1269	X	2	200017	2000110	AA	ABQ	NM	1.00	AA
4	X	ABQ	NM	1269	X	2	200017	2000111	AA	ATL	GA	1.00	AA
5	X	ATL	GA	1269	X	2	200018	2000112	AA	ABQ	NM	1.00	AA
6	X	ABQ	NM	1269	X	2	200018	2000113	AA	ATL	GA	1.00	AA
7	X	ATL	GA	1269	X	2	200019	2000114	AA	ABQ	NM	1.00	AA
8	NULL	DFW	TX	732	X	2	200019	2000115	AA	ATL	GA	1.00	AA
9	X	ABQ	NM	569	X	2	200019	2000115	AA	DFW	TX	1.00	AA
10	X	ATL	GA	1269	X	2	2000110	2000116	AA	ABQ	NM	1.00	AA
11	NULL	DFW	TX	732	X	2	2000110	2000117	AA	ATL	GA	1.00	AA
12	X	ABQ	NM	569	X	2	2000110	2000117	AA	DFW	TX	1.00	AA
13	X	ATL	GA	1269	X	2	2000111	2000118	AA	ABQ	NM	1.00	AA
14	NULL	DFW	TX	732	X	2	2000111	2000119	AA	ATL	GA	1.00	AA
15	X	ABQ	NM	569	X	2	2000111	2000119	AA	DFW	TX	1.00	AA
16	X	ATL	GA	1269	X	2	2000112	2000120	AA	ABQ	NM	1.00	AA
17	NULL	DFW	TX	732	X	2	2000112	2000121	AA	ATL	GA	1.00	AA
18	X	ABQ	NM	569	X	2	2000112	2000121	AA	DFW	TX	1.00	AA
19	X	ATL	GA	1269	X	2	2000113	2000122	AA	ABQ	NM	1.00	AA
20	NULL	DFW	TX	732	X	2	2000113	2000123	AA	ATL	GA	1.00	AA
21	X	ABQ	NM	569	X	2	2000113	2000123	AA	DFW	TX	1.00	AA
22	X	ATL	GA	1269	Y	2	2000115	2000126	AA	ABQ	NM	2.00	AA
23	NULL	ORD	IL	606	Y	2	2000115	2000127	AA	ATL	GA	2.00	AA
24	X	ABQ	NM	1118	Y	2	2000115	2000127	AA	ORD	IL	2.00	AA
25	NULL	ATL	GA	1269	X	2	2000116	2000128	AA	ABQ	NM	1.00	AA
26	X	PNS	FL	272	X	2	2000116	2000128	DL	ATL	GA	1.00	AA
27	NULL	ATL	GA	272	X	2	2000116	2000129	DL	PNS	FL	1.00	AA
28	NULL	DFW	TX	732	X	2	2000116	2000129	AA	ATL	GA	1.00	AA

Figure A.1 DB1B Coupon.

airportg	bulkfar	dest	dests	itingeot	itinid	v15	mktfar	mktid	mktmile	nonstop	op
ABQ:ATL	0	ATL	GA	2	200016	3	102	200018	1269	1269	A
ATL:ABQ	0	ABQ	NM	2	200016	3	102	200019	1269	1269	A
ABQ:ATL	0	ATL	GA	2	200017	3	107	2000110	1269	1269	A
ATL:ABQ	0	ABQ	NM	2	200017	3	107	2000111	1269	1269	A
ABQ:ATL	0	ATL	GA	2	200018	3	155	2000112	1269	1269	A
ATL:ABQ	0	ABQ	NM	2	200018	3	155	2000113	1269	1269	A
ABQ:ATL	0	ATL	GA	2	200019	3	98	2000114	1269	1269	A
ATL:DFW:ABQ	0	ABQ	NM	2	200019	3	101	2000115	1301	1269	A
ABQ:ATL	0	ATL	GA	2	2000110	3	167	2000116	1269	1269	A
ATL:DFW:ABQ	0	ABQ	NM	2	2000110	3	171	2000117	1301	1269	A
ABQ:ATL	0	ATL	GA	2	2000111	3	179	2000118	1269	1269	A
ATL:DFW:ABQ	0	ABQ	NM	2	2000111	3	184	2000119	1301	1269	A
ABQ:ATL	0	ATL	GA	2	2000112	3	202	2000120	1269	1269	A
ATL:DFW:ABQ	0	ABQ	NM	2	2000112	3	207	2000121	1301	1269	A
ABQ:ATL	0	ATL	GA	2	2000113	3	205	2000122	1269	1269	A
ATL:DFW:ABQ	0	ABQ	NM	2	2000113	3	211	2000123	1301	1269	A
ABQ:ATL:PNS	0	PNS	FL	2	2000116	4	390	2000128	1541	1172	99
PNS:ATL:DFW	0	ABQ	NM	2	2000116	4	399	2000129	1573	1172	99
ABQ:BOS:HPN	0	HPN	NY	2	2000117	5	116	2000130	2140	1830	99
HPN:JFK:DFW	0	ABQ	NM	2	2000117	4	106	2000131	1960	1830	99
ABQ:BOS:PW	0	BOS	MA	2	2000120	5	364	2000137	2164	1974	99
BOS:DFW:ABQ	0	ABQ	NM	2	2000120	5	359	2000138	2131	1974	A
ABQ:DCA	0	DCA	DC	2	2000121	4	137	2000139	1650	1650	A
ABQ:DCA	0	DCA	DC	2	2000122	4	147	2000140	1650	1650	A
ABQ:DCA	0	DCA	DC	2	2000124	4	354	2000143	1650	1650	A
DCA:DFW:ABQ	0	ABQ	NM	2	2000124	4	378	2000144	1761	1650	A
ABQ:DCA	0	DCA	DC	2	2000126	4	106	2000147	1650	1650	A
DCA:DFW:ABQ	0	ABQ	NM	2	2000126	4	114	2000148	1761	1650	A

Figure A.2 DB1B Market.

	bulkf	coupon	distance	v4	dollar	itinfare	itingeot	itinid	itinyiel	onlin	origin	originap	originst
1	0	2	2538	6	1	205	2	200016	.08	1	ABQ	1	NM
2	0	2	2538	6	1	214	2	200017	.08	1	ABQ	1	NM
3	0	2	2538	6	1	311	2	200018	.12	1	ABQ	1	NM
4	0	3	2570	6	1	199	2	200019	.08	1	ABQ	1	NM
5	0	3	2570	6	1	338	2	2000110	.13	1	ABQ	1	NM
6	0	3	2570	6	1	363	2	2000111	.14	1	ABQ	1	NM
7	0	3	2570	6	1	409	2	2000112	.16	1	ABQ	1	NM
8	0	3	2570	6	1	416	2	2000113	.16	1	ABQ	1	NM
9	0	3	2993	6	1	19	2	2000115	.01	1	ABQ	1	NM
10	0	5	3114	7	1	789	2	2000116	.25	0	ABQ	1	NM
11	0	5	4130	9	1	222	2	2000117	.05	0	ABQ	1	NM
12	0	6	5715	12	1	0	2	2000119	.00	0	ABQ	1	NM
13	0	5	4295	9	1	723	2	2000120	.17	0	ABQ	1	NM
14	0	1	1650	4	1	137	2	2000121	.08	1	ABQ	1	NM
15	0	1	1650	4	1	147	2	2000122	.09	1	ABQ	1	NM
16	0	4	3466	7	1	0	2	2000123	.00	0	ABQ	1	NM
17	0	3	3411	7	1	732	2	2000124	.21	1	ABQ	1	NM
18	0	3	3411	7	1	0	2	2000125	.00	1	ABQ	1	NM
19	0	3	3411	7	1	220	2	2000126	.06	1	ABQ	1	NM
20	0	3	3411	7	1	236	2	2000127	.07	1	ABQ	1	NM
21	0	3	3411	7	1	237	2	2000128	.07	1	ABQ	1	NM
22	0	3	3411	7	1	270	2	2000129	.08	1	ABQ	1	NM
23	0	3	3411	7	1	284	2	2000130	.08	1	ABQ	1	NM
24	0	3	3411	7	1	299	2	2000131	.09	1	ABQ	1	NM
25	0	3	3411	7	1	307	2	2000132	.09	1	ABQ	1	NM
26	0	3	3411	7	1	308	2	2000133	.09	1	ABQ	1	NM
27	0	3	3411	7	1	314	2	2000134	.09	1	ABQ	1	NM
28	0	3	3411	7	1	336	2	2000135	.10	1	ABQ	1	NM

Figure A.3 DB1B Ticket.

dest	farecla	itinid	mktid	origin	passen	v15	mktfare	opcarri	rpc	tkcarrie	v37
DFW	X	2000165	20001114	ABQ	2	2	95	AA	AA	AA	AA
DFW	X	2000166	20001115	ABQ	1	2	96	AA	AA	AA	AA
DFW	X	2000167	20001116	ABQ	1	2	97	AA	AA	AA	AA
DFW	X	2000168	20001117	ABQ	10	2	98	AA	AA	AA	AA
DFW	X	2000169	20001118	ABQ	1	2	99	AA	AA	AA	AA
DFW	X	2000170	20001119	ABQ	2	2	100	AA	AA	AA	AA
DFW	X	2000171	20001120	ABQ	1	2	108	AA	AA	AA	AA
DFW	X	2000172	20001121	ABQ	2	2	110	AA	AA	AA	AA
DFW	X	2000173	20001122	ABQ	3	2	112	AA	AA	AA	AA
DFW	X	2000174	20001123	ABQ	1	2	136	AA	AA	AA	AA
DFW	X	2000175	20001124	ABQ	1	2	162	AA	AA	AA	AA
DFW	X	2000176	20001125	ABQ	2	2	203	AA	AA	AA	AA
DFW	X	2000177	20001126	ABQ	1	2	213	AA	AA	AA	AA
DFW	X	2000178	20001127	ABQ	1	2	269	AA	AA	AA	AA
DFW	Y	2000179	20001128	ABQ	1	2	71	AA	AA	AA	AA
DFW	Y	2000180	20001129	ABQ	1	2	97	AA	AA	AA	AA
DFW	Y	2000181	20001130	ABQ	1	2	99	AA	AA	AA	AA
DFW	Y	2000182	20001131	ABQ	1	2	109	AA	AA	AA	AA
DFW	Y	2000183	20001132	ABQ	2	2	110	AA	AA	AA	AA
DFW	Y	2000184	20001133	ABQ	5	2	128	AA	AA	AA	AA
DFW	Y	2000185	20001134	ABQ	9	2	130	AA	AA	AA	AA
DFW	Y	2000186	20001135	ABQ	1	2	141	AA	AA	AA	AA
DFW	Y	2000187	20001136	ABQ	12	2	143	AA	AA	AA	AA
DFW	Y	2000188	20001137	ABQ	26	2	145	AA	AA	AA	AA
DFW	Y	2000189	20001138	ABQ	1	2	153	AA	AA	AA	AA
DFW	Y	2000190	20001139	ABQ	2	2	155	AA	AA	AA	AA
DFW	Y	2000191	20001140	ABQ	1	2	295	AA	AA	AA	AA
DFW	Y	2000192	20001141	ABQ	1	2	383	AA	AA	AA	AA

Figure A.4 DB1B Coupon and Market comined.

	farecla	itinid	dest	itinfare	itinyiel	milesflo	online	ori_t	passen	rpcarri	round
1	X	200016	ATL	205	.08	2538	1	ABQ	1	AA	1
2	X	200017	ATL	214	.08	2538	1	ABQ	1	AA	1
3	X	200018	ATL	311	.12	2538	1	ABQ	1	AA	1
4	X	200019	ATL	199	.08	2570	1	ABQ	1	AA	1
5	X	2000110	ATL	338	.13	2570	1	ABQ	1	AA	1
6	X	2000111	ATL	363	.14	2570	1	ABQ	1	AA	1
7	X	2000112	ATL	409	.16	2570	1	ABQ	1	AA	1
8	X	2000113	ATL	416	.16	2570	1	ABQ	1	AA	1
9	X	2000116	PNS	789	.25	3114	0	ABQ	1	AA	1
10	X	2000117	HPN	222	.05	4100	0	ABQ	1	AA	1
11	X	2000120	BOS	723	.17	4295	0	ABQ	1	AA	1
12	X	2000121	DCA	137	.08	1650	1	ABQ	1	AA	0
13	X	2000122	DCA	147	.09	1650	1	ABQ	1	AA	0
14	G	2000124	DCA	732	.21	3411	1	ABQ	1	AA	1
15	X	2000126	DCA	220	.06	3411	1	ABQ	1	AA	1
16	X	2000127	DCA	236	.07	3411	1	ABQ	1	AA	1
17	X	2000128	DCA	237	.07	3411	1	ABQ	1	AA	1
18	X	2000129	DCA	270	.08	3411	1	ABQ	1	AA	1
19	X	2000130	DCA	284	.08	3411	1	ABQ	5	AA	1
20	X	2000131	DCA	299	.09	3411	1	ABQ	1	AA	1
21	X	2000132	DCA	307	.09	3411	1	ABQ	1	AA	1
22	X	2000133	DCA	308	.09	3411	1	ABQ	1	AA	1
23	X	2000134	DCA	314	.09	3411	1	ABQ	1	AA	1
24	X	2000135	DCA	336	.10	3411	1	ABQ	1	AA	1
25	X	2000136	DCA	338	.10	3411	1	ABQ	1	AA	1
26	X	2000137	DCA	357	.10	3411	1	ABQ	1	AA	1
27	X	2000138	DCA	379	.11	3411	1	ABQ	1	AA	1
28	X	2000139	DCA	474	.14	3411	1	ABQ	1	AA	1

Figure A.5 DB1B Coupon,Market and Ticket comined.

A.1.2 Official Airline Guide

a) Overview:

The Official Airline Guide (OAG) provides flight schedule information. This flight information is used in calculating the Herfindahl Index in the analysis.The OAG provides information about the frequency of airlines and the seats offered by individual airlines between 419 x 419 airport pair's. Figure A.6 shows the description of the fields provided by the OAG.

carrier	carr1nam	depai	depsta	arrairpo	arrsta	seats	fstseat	bussea	ecosea	gena	gnacft
AA	American Airlines	SAN	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SAN	CA	SBA	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SAN	CA	SBA	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SAN	CA	SBA	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SAN	CA	SBA	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBA	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBP	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBP	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBP	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBP	CA	LAX	CA	34	0	0	34	SF3	Saab SF
AA	American Airlines	SBP	CA	LAX	CA	34	0	0	34	SF3	Saab SF

Figure A.6 Official Airline Guide.

A.2 Source Code

<u>Code Name</u>: “fare_twoway_oneway1.m”

Input Files:

- airporlist_ID.mat -This file is the list of 419 airports.
- “od_cmt_2000.dat” - This file is the main input data file. This is the combined Coupon,Market and Ticket file.

Functions called:

- “find_airport.m” - This function determines the location of the airport in the airport list.

Output Files:

- roundtrip_bus_od_2000.mat - 419 x 419 table of Histogram of roundtrip business fares.
- roundtrip_nonbus_od_2000.mat-419 x 419 table of Histogram of roundtrip coach fares.
- oneway_bus_od_2000.mat- 419 x 419 table of Histogram of oneway business fares.
- oneway_nonbus_od_2000.mat- 419 x 419 table of Histogram of oneway business fares.
- pax_rt_b_2000.mat-419 x 419 table of Roundtrip business passenger flows.
- pax_rt_nb_2000.mat- 419 x 419 table of Roundtrip coach passenger flows.
- pax_oneway_b_2000.mat- 419 x 419 table of Oneway business passenger flows.
- pax_oneway_nb_2000.mat-419 x 419 table of Oneway coach passenger flows.

Sample Output file deatils:

```
roundtrip_nonbus_od_2000_1(21,32).hist

ans =

Columns 1 through 12

0    0    4    5   10   22    6    2    6    7    9    6

Columns 13 through 24

10    8   13    8    6   12   20   10    9    1    3    0

Columns 25 through 36

1    0    1    1    0    0    0    0    0    0    0    0

Columns 37 through 48
```

0 0 0 0 0 0 0 0 0 0 0 0

Columns 49 through 60

0 0 0 0 0 0 0 0 0 0 0 0

Columns 61 through 72

0 0 0 0 0 0 0 0 0 0 0 0

Columns 73 through 84

0 0 0 0 0 0 0 0 0 0 0 0

Columns 85 through 96

0 0 0 0 0 0 0 0 0 0 0 0

Columns 97 through 100

0 0 0 0

Code deatils:

```
% This programme calculates the fare per mile for both business and non
% business fares . They are spilt into 2 categories which are as follows:
% CATEGORY 1:
% 1)Nonbusiness fares include only X and class fareS
% 2)Y,G,Fand N are business class fares.
% CATEGORY 2:
% 1)Nonbusiness fares include only X and Y class fares.
% 2)G,Fand N are business class fares.
% All fares below $ 50 and above $4000 are deleted.
clear all;
clc
load airportlist_ID;
% Intializing the data set
for i = 1:419
```

```
for j = 1:419
roundtrip_bus_od_2000(i,j).hist = zeros(1,100);
roundtrip_nonbus_od_2000(i,j).hist   = zeros(1,100);
oneway_bus_od_2000(i,j).hist      = zeros(1,100);
oneway_nonbus_od_2000(i,j).hist   = zeros(1,100);
end
end
pax_rt_b_2000(1:419,1:419) =0;
pax_rt_nb_2000(1:419,1:419) =0;
pax_oneway_b_2000(1:419,1:419) =0;
pax_oneway_nb_2000(1:419,1:419) =0;
fid = fopen('od_cmt_2000.dat','r'); % Input File
if(fid < 0)
else
   % skip the first line
    firstLine = fgets(fid);
end
line     = '';
c   = 0;
% Intialising the Bins
 bins_business    = 0.05:0.05:5;
bins_nonbusiness = 0.05:0.05:5;
while ~ feof(fid)
if fix(c/50000)==c/50000
         disp(c);
      end

line                 = fgets(fid);
length_line          = length(line);
```

```
double_line         = double(line);
[strt_line,end_line] = size(double_line);
blanks_position      = find(double_line==9);
fare_class           = line(1:blanks_position(1)-1);  % Reading the fare class
roundtripindicator    = str2num(line(blanks_position(9):end_line)); % eading roundtrip indicator
if isempty(roundtripindicator)  % Checks for empty cells
      continue
   end
dest       = line(blanks_position(2)+1:blanks_position(3)-1); % Reads the destination
ori        = line(blanks_position(7)+1:blanks_position(8)-1); % Reads the origin
itfare     = str2num(line(blanks_position(3):blanks_position(4))); % Reads the itineary fare
ityield    = str2num(line(blanks_position(4):blanks_position(5))); % Reads the fare per mile
%     online     = str2num(line(blanks_position(7):blanks_position(8)));
pas        = str2num(line(blanks_position(8):blanks_position(9))); % Reads for passenger
if (double(ori) == 32 | double(dest) == 32 )
      continue
   end
origin         = find_airport(ori,airportlist_ID);  % Determines the origin airport in the O-D table
destination    = find_airport(dest,airportlist_ID); % Determines the origin airport in the O-D table
if (isempty(origin)| isempty(destination))  % Checks if the cells are empty
      continue
   end
% if ( itfare >= 50 & itfare <= 4000 ) % Checks for fares
if (roundtripindicator == 1)    % Checks for round trip fare
if (fare_class=='C'|fare_class=='D'|fare_class=='F'|fare_class=='G')  %  C,D,F and G  form the business class
n_bus = hist(ityield , bins_business);
roundtrip_bus_od_2000(origin,destination).hist = roundtrip_bus_od_2000(origin,destination).hist + n_bus ;  % Creats the bins i.e. histogram of fare per mile
pax_rt_b_2000(origin,destination) = pax_rt_b_2000(origin,destination) + pas ; % Determines the passenger flows.
```

```
elseif (fare_class =='X'|fare_class=='Y') % X and Y form the Nonbusiness Class
n_nonbus = hist( ityield , bins_nonbusiness);
roundtrip_nonbus_od_2000(origin,destination).hist = roundtrip_nonbus_od_2000(origin,destination).hist + n_nonbus; % Creats the bins i.e. histogram of fare per mile
pax_rt_nb_2000(origin,destination) = pax_rt_nb_2000(origin,destination) + pas ; % Determines the passenger flows.
        end
elseif (roundtripindicator == 0) % Checks for Oneway fare
if (fare_class=='C'|fare_class=='D'|fare_class=='F'|fare_class=='G')  %  C,D,F and G  form the business class
n_bus = hist(ityield , bins_business);
oneway_bus_od_2000(origin,destination).hist = oneway_bus_od_2000(origin,destination).hist + n_bus ;  % Creats the bins i.e. histogram of fare per mile
pax_oneway_b_2000(origin,destination) = pax_oneway_b_2000(origin,destination) + pas ;   % Determines the passenger flows.
elseif (fare_class =='X'|fare_class=='Y') % X and Y form the Nonbusiness Class
n_nonbus = hist( ityield , bins_nonbusiness);
oneway_nonbus_od_2000(origin,destination).hist=oneway_nonbus_od_2000(origin,destination).hist + n_nonbus; % Creats the bins i.e. histogram of fare per mile
pax_oneway_nb_2000(origin,destination = pax_oneway_nb_2000(origin,destination) + pas ; % Determines the passenger flows.
        end
      end
  end

  c = c + 1;
end

fclose(fid);

     % save workspace,
--------------------------------------------------------------------------------------------------------------------------------------------------------------
```

<u>Code Name</u>: "find_airport.m"

Input Files:

- airporlist_ID.mat -This file is the list of 419 airports.

Output :

- Location of the airport in the airport list e.g. "29"

Code deatils:

```
% This function determines the airport number from the airport list.

function arprt =find_airport(chr,airports_419);

numChar = size(airports_ID,1);

[m,n]=size(airports_ID);

num_airport=zeros(m,1);

x=double(chr);

asc=10000*x(1)+100*x(2)+x(3); % Converting to ASCII values

for i=1:numChar

    y=double(airports_ID(i,:));

    num_airport(i)=10000*y(1)+100*y(2)+y(3);

end

arprt =find(num_airport==asc);

return
```

<u>Code Name</u>: "fpm_std_pax_sum.m"

Input Files:

- roundtrip_bus_od_2000.mat -419 x 419 table of Histogram of roundtrip business fares.
- roundtrip_nonbus_od_2000.mat-419 x 419 table of Histogram of roundtrip coach fares.
- pax_rt_b_2000.mat-419 x 419 table of Roundtrip business passenger flows.
- pax_rt_nb_2000.mat-419 x 419 table of Roundtrip coach passenger flows.

Output Files:

- fpm_rt_bus_od_2000.mat - 419 x 419 table of Fare per mile of roundtrip business fares.

- fpm_rt_nonbus_od_2000.mat - 419 x 419 table of Fare per mile of roundtrip coach fares.
- std_rt_bus_od_2000.mat -419 x 419 table of Standard deviation of roundtrip business fares.
- std_rt_nonbus_od_2000.mat -419 x 419 table of Standard deviation of roundtrip coach fares.

Code deatils:

```
% This programe calculates the fare per mile , standard deviation of the
% fare per mile , total sum of records , total passenger flows. This data
% is related to round trip passenger flows.
clear all ;
clc;
% %-------------------------------------------------------------------------
% % This section of the program adds the histogram for all quarters to make a
% % data set for the entire year 2000.
load roundtrip_bus_od_2001_1;
load roundtrip_bus_od_2001_2;
load roundtrip_bus_od_2001_3;
load roundtrip_bus_od_2001_4;
Intialize matrix.
for i = 1:419
for j = 1:419
roundtrip_bus_od_2001(i,j).hist = zeros(1,100);
roundtrip_nonbus_od_2001(i,j).hist = zeros(1,100);
end
end
load roundtrip_nonbus_od_2001_1;
load roundtrip_nonbus_od_2001_2;
load roundtrip_nonbus_od_2001_3;
load roundtrip_nonbus_od_2001_4;
% Determing the bins for the enitire yar of 2000 by adding the bins for all the quarters
for i = 1:419
```

```
 for j = 1:419
xb_1   = roundtrip_bus_od_2001_1(i,j).hist;
xb_2   = roundtrip_bus_od_2001_2(i,j).hist;
xb_3   = roundtrip_bus_od_2001_3(i,j).hist;
xb_4   = roundtrip_bus_od_2001_4(i,j).hist;
xnb_1   = roundtrip_nonbus_od_2001_1(i,j).hist;
xnb_2   = roundtrip_nonbus_od_2001_2(i,j).hist;
xnb_3   = roundtrip_nonbus_od_2001_3(i,j).hist;
xnb_4   = roundtrip_nonbus_od_2001_4(i,j).hist;
xb    = xb_1 + xb_2 + xb_3 + xb_4 ;
xnb   = xnb_1 + xnb_2 + xnb_3 + xnb_4  ;
% Determing the bins for the enitire yar of 2001 by adding the bins for all the quarters
roundtrip_bus_od_2001(i,j).hist    = roundtrip_bus_od_2001(i,j).hist + xb;
roundtrip_nonbus_od_2001(i,j).hist = roundtrip_nonbus_od_2001(i,j).hist + xnb;
end
end
% %-------------------------------------------------------------------------
% % This part calculates the mean fare per mile value , standard deviation
% % and sum of records between each o-d pair.
load roundtrip_bus_od_2001;
load roundtrip_nonbus_od_2001;
x_bus   = [ 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 1.05  1.1 1.15 1.2 1.25 1.3 1.35 1.4 1.45 1.5 1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95 2 2.05 2.1 2.15 2.2 2.25 2.3 2.35 2.4 2.45 2.5 2.55 2.6 2.65 2.7 2.75 2.8 2.85 2.9 2.95 3 3.05 3.1 3.15 3.2 3.25 3.3 3.35 3.4 3.45 3.5 3.55 3.6 3.65 3.7 3.75 3.8 3.85 3.9 3.95 4 4.05 4.1 4.15 4.2 4.25 4.3 4.35 4.4 4.45 4.5 4.55 4.6 4.65 4.7 4.75 4.8 4.85 4.9 4.95 5];
x_nonbus = [ 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75 0.8 0.85 0.9 0.95 1 1.05  1.1 1.15 1.2 1.25 1.3 1.35 1.4 1.45 1.5 1.55 1.6 1.65 1.7 1.75 1.8 1.85 1.9 1.95 2 2.05 2.1 2.15 2.2 2.25 2.3 2.35 2.4 2.45 2.5 2.55 2.6 2.65 2.7 2.75 2.8 2.85 2.9 2.95 3 3.05  3.1 3.15 3.2 3.25 3.3 3.35 3.4 3.45 3.5 3.55 3.6 3.65 3.7 3.75 3.8 3.85 3.9 3.95 4 4.05 4.1 4.15 4.2 4.25 4.3 4.35 4.4 4.45 4.5 4.55 4.6 4.65 4.7 4.75 4.8 4.85 4.9 4.95 5];
sum_bus_od_2001(1:419,1:419)              = 0;
```

```
sum_nonbus_od_2001(1:419,1:419)          = 0;
fpm_rt_bus_od_2001(1:419,1:419)          = 0;
fpm_rt_nonbus_od_2001(1:419,1:419)       = 0;
stdeviation_rt_bus_od_2001(1:419,1:419)    = 0;
stdeviation_rt_nonbus_od_2001(1:419,1:419) = 0;
for i = 1:419
  for j = 1:419
total_bus    = sum(roundtrip_bus_od_2001(i,j).hist);
total_nonbus = sum(roundtrip_nonbus_od_2001(i,j).hist);
sum_bus_od_2001(i,j)    = total_bus;       % Total Business records between the o-d pairs
sum_nonbus_od_2001(i,j) = total_nonbus;    % Total Business records between the o-d pairs
if ( total_bus > 0 )
      prob_bus = (roundtrip_bus_od_2001(i,j).hist / total_bus);
    else
      prob_bus = 0;
    end
product_bus            = (x_bus .* prob_bus);
fpm_rt_bus_od_2001(i,j) = sum(product_bus); % Fare per Mile
sqr_x_bus              = (x_bus).^2;
pro_std_bus            = (sqr_x_bus.* prob_bus);
sum_pro_std_bus        =  sum( pro_std_bus );
stdeviation_rt_bus_od_2001(i,j) = sqrt(abs(sum_pro_std_bus - fpm_rt_bus_od_2001(i,j))); % Standard Deviation
if ( total_nonbus > 0 )
      prob_nonbus = (roundtrip_nonbus_od_2001(i,j).hist / total_nonbus);
    else
      prob_nonbus = 0;
    end
product_nonbus            = (x_nonbus .* prob_nonbus);
fpm_rt_nonbus_od_2001(i,j) =  sum(product_nonbus);   % Fare per Mile
```

```
sqr_x_nonbus              = (x_nonbus).^2;
pro_std_nonbus            = (sqr_x_nonbus.* prob_nonbus);
sum_pro_std_nonbus        =  sum( pro_std_nonbus );
stdeviation_rt_nonbus_od_2001(i,j) = sqrt(abs(sum_pro_std_nonbus - fpm_rt_nonbus_od_2001(i,j))); % Stdandard Deviation
end
end
```

Code Name: "GCD.m"

Input Files:

- airportlist_419.mat-This is a mat file containing information about airport latitude and longitude.

Output File:

- interAirportDist.mat - 419 x 419 table of Great Circle distance between airports in statue miles.

Code details:

```
% This is a program to Fares and GCD between  O-D pairs in Statue Miles
clear all;
load airportlist_419; %includes airport
[m,n] = size(airportlist_419);
% Calculate the GCD between the 2 O-D pairs
for i=1:n
   for j=1:n
         interAirportDist(i,j)=deg2sm(distance(airportlist_419(i).lat,airportlist_419(i).lon,airportlist_419(j).lat,airportlist_419(j).lon)); %
Great circle distance
   end % for i=1:n
end % for j=1:n
```

<u>Code Name</u>: “find_airline_oag.m”

Input Files:

- airlines_oag.mat -This file is the list of all the airlines in OAG.

Output :

- Location of the airline in the airline list e.g. “3”

Code deatils:

```
% This function determines the airline
function airline  = find_airline_oag(airl_oag,airlines_oag)
total_airlines = size(airlines_oag,1);
[m,n]          = size(airlines_oag); % List of all the airlines in OAG.
num_airline    = zeros(m,1);
x              = double(airl_oag);
asc            = 100*x(1)+x(2);
for i=1:total_airlines
   y            = double(airlines_oag(i,:));
   num_airline(i)= 100*y(1)+ y(2);
end
airline          = find(num_airline==asc);
```

<u>Code Name</u>: “find_aircraft_oag.m”

Input Files:

- aircraft_oag.mat -This file is the list of all the aircraft in OAG.

Output :

- Location of the aircraft in the aircraft list e.g. “5”

Code deatils:

```
% This functions determines the type of aircraft.
function aircraft  = find_aircraft_oag(aircft_oag,aircrafts_oag)
```

```
total_aircrafts = size(aircrafts_oag,1); % 'aircrafts_oag.mat is a file where the list of all the aircrafts in OAG exists.
[m,n]          = size(aircrafts_oag);
num_aircraft    = zeros(m,1);
x              = double(aircft_oag);
asc            = 100000*x(2)+1000*x(3)+x(4);
for i=1:total_aircrafts
    y           = double(aircrafts_oag(i,:));
   num_aircraft(i)= 100000*y(1)+1000*y(2)+y(3);
end
aircraft        = find(num_aircraft==asc);
```

<u>Code Name</u>: "frequency_seats_airline_aircraft_oag.m"

Input Files:

- airports_ID-This file is the list of 419 airports.
- airlines_oag- This file contains the list of all airlines in the OAG.
- aircrafts_oag - This file contains the list of all aircrafts in the OAG.
- oag_2000_1.dat- This is the main input data file containing information of flights for the year 2000.

Functions called:

- "find_airport.m" - This function determines the location of the airport in the airport list.
- "find_aircraft_oag.m"- This function determines the location of the aircraft in the aircraft list.
- "find_airline_oag.m"- This function determines the location of the airline in the airline list.

Output File:

- airlines_aircrafts_frequency.mat-This contains 419 x 419 table of the frequency of different aircrafts used by airlines for all O-D pairs.
- oag_2001_1_freq.mat - This mat file has 419 x 419 table of the frequecny of all airlines between the O-D pairs.
- oag_2001_1_arl_arcft.mat - This mat file has 419 x 419 table of the frequency between all O-D having individual value of frequency for airline and the aircraft used by the airline.

- oag_2001_1_t_seats.mat - This mat file has 419 x 419 table of the total number of seats flow between all O-D having individual value of seats for airline and the aircraft used by the airline.
- oag_2001_1_tseats.mat - This mat file has 419 x 419 table of the total number of seats of all airlines between the O-D

Code details:

```
% This programme calculates the Frequecy of flights for the year 2001 -
% quarter 1. This programme uses data from OAG having both Op flights
% This programe calls the following functions:
% 1) find_airline_oag.m
% 2) find_aircraft_oag.m
% 3) find_airport
% The input data for this file is 'oag_2000_1.dat'
clear all;
load airports_ID;
load airlines_oag;
load aircrafts_oag;
load oag_2001_1_freq;
%-------------------------------------------------------------------
% Initialising for calculating the frequency of the type of aircraft
% used by airlines                    USED : Below in PART 1
airlines_aircrafts_frequency(48,60)=0;
%-------------------------------------------------------------------
% Initialising for frequency          USED : Below in PART 2
oag_2001_1_freq(419,419)=0;
%-----------------------------------------------------------------
% Intialising for the O-D matrix      USED : Below in PART 3
[o,d,freq] = find(oag_2001_1_freq);
[r,s] = size(o);
for k = 1:r
```

```
    or = o(k);
    dt = d(k);
    for p = 1:48     % The total number of airlines in OAG
        for q = 1:60 % The total number of aircraft in OAG
            oag_2001_1_arl_arcft(or,dt).freq(p,q)=0;
end
end
end
%-------------------------------------------------------------------
% Intialising for the O-D matrix   USED : Below in PART 4
[o,d,freq] = find(oag_2001_1_freq);
[r,s] = size(o);
for k = 1:r
    or = o(k);
    dt = d(k);
    for p = 1:48     % The total number of airlines in OAG
        for q = 1:60 % The total number of aircraft in OAG
            oag_2001_1_t_seats(or,dt).freq(p,q)=0;
        end
    end
end
%-------------------------------------------------------------------
% Intialising for the O-D matrix   USED : Below in PART 5
[o,d,freq] = find(oag_2001_1_freq);
[r,s] = size(o);
for k = 1:r
    or = o(k);
    dt = d(k);
    for p = 1:48     % The total number of airlines in OAG
```

```
        for q = 1:60 % The total number of aircraft in OAG
            oag_2001_1_t_fbseats(or,dt).freq(p,q)=0;
        end
    end
end
%------------------------------------------------------------------
% Intialising for the O-D matrix   USED : Below in PART 6
[o,d,freq] = find(oag_2001_1_freq);
[r,s] = size(o);
for k = 1:r
    or = o(k);
    dt = d(k);
    for p = 1:48     % The total number of airlines in OAG
        for q = 1:60 % The total number of aircraft in OAG
            oag_2001_1_t_eseats(or,dt).freq(p,q)=0;
        end
    end
end
%------------------------------------------------------------------
% Initialising for frequency           USED : Below in PART 7
oag_2001_1_tseats(419,419)=0;
%------------------------------------------------------------------
% Initialising for calculating the total number of the type of aircraft
% used by airlines                     USED : Below in PART 8
airlines_aircrafts_tseats(48,60)=0;
%------------------------------------------------------------------
% Initialising for frequency           USED : Below in PART 9
oag_2001_1_fbseats(419,419)=0;
%------------------------------------------------------------------
```

```
% This section opens the file

fid = fopen('oag_2001_1.dat','r');
line     = '';
while ~feof(fid)
line     = fgets(fid);
length_line = length(line);
double_line = double(line);
[strt_line,end_line] = size(double_line);
blanks_position=find(double_line==9);
airline = line(1:blanks_position(1));
x  = double(airline);
s  = sum(x);
if ( s == 744 )
        continue
    end
aircraft = line(blanks_position(3):blanks_position(4));  % Aircraft
frequency = str2num(line(blanks_position(4):blanks_position(5))); % frequenct of flight
origin = line(blanks_position(1):blanks_position(2)); % Origin
dest   = line(blanks_position(2):blanks_position(3)); % Destination
ori    = find_airport(origin,airports_ID);% Determines the location of origin in the O-D table
des    = find_airport(dest,airports_ID);% Determines the location of destination in the O-D table
arcft  = find_aircraft_oag(aircraft,aircrafts_oag);% Determines the aircraft
arl    = find_airline_oag(airline,airlines_oag);% Determines the airline
t_seats   = str2num(line(blanks_position(5):blanks_position(6)));% This is the total number of seats
t_eseats = str2num(line(blanks_position(6):blanks_position(7)))
t_fbseats = str2num(line(blanks_position(7):end_line));% This is the total number of business and first class seats
if (isempty(ori) | isempty(des)) % checks whether origin and destination is empty
        continue
```

```
    end
    %--------------------------------------------------------------------
    % PART 1
    % This section calculates the frequency of different aircrafts used by
    % airlines for all O-D pairs
airlines_aircrafts_frequency(arl,arcft)= frequency + airlines_aircrafts_frequency(arl,arcft);
%--------------------------------------------------------------------
    % PART 2
    % This section calculates the frequecny of all airlines between the O-D
    % pairs
oag_2001_1_freq(ori,des) = frequency + oag_2001_1_freq(ori,des);

    %--------------------------------------------------------------------
    % PART 3
    % This section calculates the frequency between all O-D having
    % individual value of frequency for airline and the aircraft used by
    % the airline.This will be the input for caculating the Herfidahl index
oag_2001_1_arl_arcft(ori,des).freq(arl,arcft) = frequency + oag_2001_1_arl_arcft(ori,des).freq(arl,arcft);

%--------------------------------------------------------------------
    % PART 4
    % This section calculates the total number of seats flow between all O-D having
    % individual value of seats for airline and the aircraft used by
    % the airline.This will be the input for caculating the Herfidahl index

oag_2001_1_t_seats(ori,des).freq(arl,arcft) = t_seats + oag_2001_1_t_seats(ori,des).freq(arl,arcft);

    %--------------------------------------------------------------------
    % PART 5
```

```
   % This section calculates the total number of first and business seats flow between all O-D having
   % individual value of first and business seats for airline and the aircraft used by
   % the airline.This will be the input for caculating the Herfidahl index
oag_2001_1_t_fbseats(ori,des).freq(arl,arcft) = t_fbseats + oag_2001_1_t_fbseats(ori,des).freq(arl,arcft);
%--------------------------------------------------------------------
   % PART 6
   % This section calculates the total number of coach seats flow between all O-D having
   % individual value of coach seats for airline and the aircraft used by
   % the airline.This will be the input for caculating the Herfidahl index
oag_2001_1_t_eseats(ori,des).freq(arl,arcft) = t_eseats + oag_2001_1_t_eseats(ori,des).freq(arl,arcft);
%--------------------------------------------------------------------
   % PART 7
   % This section calculates the total number of seats of all airlines between the O-D
   % pairs
oag_2001_1_tseats(ori,des) = t_seats + oag_2001_1_tseats(ori,des);
%--------------------------------------------------------------------
   % PART 8
   % This section calculates the total number  of different aircrafts used by
   % airlines for all O-D pairs
airlines_aircrafts_tseats(arl,arcft)= t_seats + airlines_aircrafts_tseats(arl,arcft);
   %--------------------------------------------------------------------
   % PART 9
   % This section calculates the total number of first and business seats of all airlines between the O-D
   % pairs
oag_2001_1_fbseats(ori,des) = t_fbseats + oag_2001_1_fbseats(ori,des);
%--------------------------------------------------------------------
end
fclose(fid);
```

Code Name: “fare_determine.m”

Input Files:

- Origin Airport Code.
- Destination Airport Code.
- The type of fare class ‘C’ and ‘B’.

Output :

- The fare values e.g.“510”

Code deatils:

```
% This function calculates the fare price between any Origin  and
% Destination.The following are the input parameters :
% 1) Origin Airport Code.
% 2) Destination Airport code.
% 3) The Type of Fare Class for Coach 'C' and Business 'B' respectively
function fare = fare_determine(ori,dest)
% ---------------------------------------------------------------------------------------------------------------------
    % Deter mining the Origin and Destination Airports and the
    % corresponding airport types

    % Loading the list of airports
    load airports_ID;
    load airportlist_419;

    origin      = ori ;
    destination  = dest ;
% Checking for airport codes
    origin_code      = find_airport(origin,airports_ID);
    destination_code = find_airport(destination,airports_ID);
```

```
    if (isempty(origin_code))
        fprintf('You have entered a wrong origin airport code \n' );
        return
    elseif(isempty(destination_code))
        fprintf('You have entered a wrong destination airport code \n' );
        return
    end
origin_type = airportlist_419(origin_code).type;
dest_type   = airportlist_419(destination_code).type;
% -----------------------------------------------------------------------------------------------------------------------
% This part takes the input whether the class is Coach or Business
fare_class  = input('Enter the fare class , if Coach then type "C" if Business type then type "B" : ' );
% -----------------------------------------------------------------------------------------------------------------------
% For Coach Class  we calculate the fare in ($).
if (upper(fare_class) == 'C' )
% Checking for whether the flight is direct flight or indirect flight.
   load herfindahl_index_avseat;   % These are herfindahl indices for direct flights
   load hindex_avseat_od;          % These are herfindahl indices for indirect flights calculated by the " Weighted Methodology
h_direct   = herfindahl_index_avseat(origin_code,destination_code);
   h_indirect = hindex_avseat_od(origin_code,destination_code);
% Checks for empty values
   if ( h_indirect == 0 & h_direct == 0 )
       fprintf('There is no flight between these 2 O-D pairs , please enter a new O-D pair \n');
       return
   end
% Loading the interAirportDist , Passenger flow matrix determining
  % whether the airport has low cost airline presence.
    load interAirportDist;         % This is the inter Airport distance matrix
```

```
    load pax_2000_1;               % This is the passenger flow matrix
    load lowcost_od;               % This is the low cost airline fare per mile values
lc_od    = lowcost_od(origin_code,destination_code);
    distance = interAirportDist(origin_code,destination_code);
    pass     = pax_2000_1(origin_code,destination_code);
if (lc_od == 0 )
        low_cost = 0;
    else
        low_cost = 1;
end
% Checks for empty values if not calculates the fare value.
if( h_direct == 0)
       fprintf('There is no direct flight between these O-D pair  \n');
       fare_coach_d = 0;
   elseif(h_direct ~= 0)
          fare_coach_d = (0.171  + (106.29 / distance )+ 0.02*(h_indirect) - 7.8e-7*(pass) - 0.105*(low_cost) + 0.009*(origin_type) + 0.004*(dest_type))*distance;
        fprintf('The value of coach fare for the given  O-D pair is : %4.2f \n',fare_coach_d);
   end
if( h_indirect == 0)
       fare_coach_w = 0 ;
       fprintf('There is no indirect flight between these O-D pairs  \n');
   elseif(h_indirect ~= 0)
         fare_coach_w  = (0.039  + (150.69 / distance )+ 0.006*(h_direct) - 2.9e-7*(pass) - 0.059*(low_cost) + 0.009*(origin_type) + 0.012*(dest_type))*distance;
       fprintf('The value of coach fare for the given  O-D pair by weighted method is : %4.2f \n',fare_coach_w);
   end
if (fare_coach_w ~= 0 & fare_coach_d ~= 0)
       fare = [fare_coach_d , fare_coach_w];
   elseif (fare_coach_w ~= 0)
```

```
      fare = fare_coach_w;
   elseif (fare_coach_d ~= 0)
      fare = fare_coach_d;
   end
% ------------------------------------------------------------------------------------------------------------------------
% For Business Class  we calculate the fare in ($).
elseif (upper(fare_class) == 'B' )

   % Loading the interAirportDist , Passenger flow matrix determining
load interAirportDist;         % This is the inter Airport distance matrix
load pax_bus_2000_1;           % This is the passenger flow matrix
distance   = interAirportDist(origin_code,destination_code);
pass       = pax_bus_2000_1(origin_code,destination_code);
% Checks for empty values if not calculates the fare value.
    if ( pass == 0 )
       fprintf('There is no flow of business passengers between these two O-D airports ');
       return
    end
if ( pass ~= 0)
fare_bus_d = (0.514 + (106.29 / distance ) - 1.2e-5*(pass) - 0.085*(origin_type) - 0.087*(dest_type))*distance;
fprintf('The value of business class fare the given  O-D pair is : %4.2f \n',fare_bus_d);
fare = fare_bus_d;
end
end
% ------------------------------------------------------------------------------------------------------------------------
```

A.3 SPSS SYNTAX

```
*--------------------------------------READ ME--------------------------------------------------------------.
* This data anylsis procedure is performed on airline fare databses (DB1B) for the year 2004.
* This procedure requires the following input databases for the year 2004 (Includes 4 Quarters):Coupon , Market & Ticket
* This syntax procedure requires a certain file and folder set up as explained in DB1B_File_Folder_Setup.doc.
* This procedure requires a File Handle as a critical input that associates with the location of the Folder
* with the respective drive.
* This procedure develops a critical input file used in the Fare Model.('od_cmt_2004.sav').
* This syntax was created at Virginia Tech by Krishna Murthy for SATS Lab.
*--------------------------------------READ ME--------------------------------------------------------------.
*-----------------------------------------------------------------------------------------------------------------------.
* DATABASE = TICKET YEAR = 2004   QUARETER = 1.
*-----------------------------------------------------------------------------------------------------------------------.
*  Import Ticket data for Quarter 1, Delete unwanted fields & Save file as a 'od_t_XXXX_1.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------------.
*-----------------------------------------------------------------------------------------------------------------------.
*      Assign appropriate file handle.
*-----------------------------------------------------------------------------------------------------------------------.

FILE HANDLE AssignedDir /Name = 'D:\DB1B'.
*-----------------------------------------------------------------------------------------------------------------------.
GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_1\Ticket\Origin_and_Destination_Survey_DB1BTicket_2004_1.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
```

```
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 Coupons F1.0
 Year F4.2
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A10
 OriginWac F2.1
 RoundTrip F4.2
 OnLine F4.2
 DollarCred F1.0
 ItinYield F6.2
 RPCarrier A2
 Passengers F5.2
 ItinFare F7.2
 BulkFare F4.2
 Distance F7.2
 DistanceGroup F2.1
 MilesFlown F7.2
 ItinGeoType F1.0
 V25 F1.0
 .
EXECUTE.
```

```
*------------------------------------------------------------------------------------------------------------------.
*  Delete unwanted variables.
*------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac OnLine DollarCred BulkFare DistanceGroup MilesFlown ItinGeoType V25 .

*------------------------------------------------------------------------------------------------------------------.
* Sort Cases by It ID.
*------------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*------------------------------------------------------------------------------------------------------------------.
* Delete all fares less than $ 50 & greater than $ 5000
*------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 50 & ItinFare <= 5000).
EXECUTE .

*------------------------------------------------------------------------------------------------------------------.
* Restrict Distance from 100 miles to  10000 miles
*------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
```

```
SELECT IF(Distance >= 100 & Distance <= 10000).

EXECUTE .

*-----------------------------------------------------------------------------------------------------------------.

*  Save file as a 'od_t_XXXX_1.sav' file where XXXX = year.

*-----------------------------------------------------------------------------------------------------------------.

SAVE OUTFILE='AssignedDir\2004\2004_1\Ticket\od_t_2004_1.sav'

 /COMPRESSED.

*-------------------------------------------------------------------------------------------------------------------------.


* DATABASE = Market YEAR = 2004   QUARETER = 1.

*-----------------------------------------------------------------------------------------------------------------.

*  Import Market data for Quarter 1, Delete unwanted fields & Save file as a 'od_m_XXXX_1.sav' file where XXXX = year.

*-----------------------------------------------------------------------------------------------------------------.


GET DATA  /TYPE = TXT

 /FILE = 'AssignedDir\2004\2004_1\Market\Origin_and_Destination_Survey_DB1BMarket_2004_1.csv'

 /DELCASE = LINE

 /DELIMITERS = ","

 /QUALIFIER = '"'

 /ARRANGEMENT = DELIMITED

 /FIRSTCASE = 2

 /IMPORTCASE = ALL

 /VARIABLES =

 ItinID F12.0

 MktID F12.0

 MktCoupons F1.0

 Year F4.2

 Quarter F1.0

 Origin A3
```

OriginAptInd F1.0

OriginCityNum F5.2

OriginCountry A2

OriginStateFips F2.1

OriginState A2

OriginStateName A14

OriginWac F2.1

Dest A3

DestAptInd F1.0

DestCityNum F5.2

DestCountry A2

DestStateFips F2.1

DestState A2

DestStateName A20

DestWac F2.1

AirportGroup A19

WacGroup A14

TkCarrierChange F4.2

TkCarrierGroup A11

OpCarrierChange F4.2

OpCarrierGroup A11

RPCarrier A2

TkCarrier A2

OpCarrier F2.1

BulkFare F4.2

Passengers F4.2

MktFare F7.2

MktDistance F7.2

MktDistanceGroup F2.1

```
 MktMilesFlown F7.2
 NonStopMiles F7.2
 ItinGeoType F1.0
 MktGeoType F1.0
 V40 F1.0
 .
EXECUTE.

*------------------------------------------------------------------------------------------------------------------.
*  Delete insignificant variables
*------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES MktCoupons OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac  WacGroup TkCarrier  MktDistanceGroup
 TkCarrierChange  TkCarrierGroup  OpCarrierChange  OpCarrierGroup  OpCarrier Passengers ItinGeoType BulkFare MktGeoType V40 MktMilesFlown
  NonStopMiles.

*------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_m_XXXX_1.sav' file where XXXX = year.
*------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_1\Market\od_m_2004_1.sav'
 /COMPRESSED.
```

```
*------------------------------------------------------------------------------------------------------------------------.
* DATABASE = Coupon YEAR = 2004   QUARETER = 1.
*------------------------------------------------------------------------------------------------------------------.
*  Import Coupon data for Quarter 1, Delete unwanted fields & Save file as a 'od_c_XXXX_1.sav' file where XXXX = year.
*------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_1\Coupon\Origin_and_Destination_Survey_DB1BCoupon_2004_1.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 MktID F12.0
 SeqNum F1.0
 Coupons F1.0
 Year F4.0
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A12
 OriginWac F2.1
```

```
 Dest A3
 DestAptInd F1.0
 DestCityNum F5.2
 DestCountry A2
 DestStateFips F2.1
 DestState A2
 DestStateName A12
 DestWac F2.1
 Break A1
 CouponType A1
 TkCarrier A2
 OpCarrier A2
 RPCarrier A2
 Passengers F4.0
 FareClass A1
 Distance F7.2
 DistanceGroup F1.0
 Gateway F4.2
 ItinGeoType F1.0
 CouponGeoType F1.0
 V35 F1.0
 .
EXECUTE.

*-------------------------------------------------------------------------------------------------------------------------.
*  DELETE UNWANTED VARIABLES
*-------------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
```

```
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac
 Break TkCarrier OpCarrier RPCarrier DistanceGroup Gateway ItinGeoType CouponGeoType V35.
*-----------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*-----------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*-----------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_c_XXXX_1.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_1\Coupon\od_c_2004_1.sav'
 /COMPRESSED.



*-----------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON & MARKET USING MARKET ID
*-----------------------------------------------------------------------------------------------------------------------.

GET
 FILE='AssignedDir\2004\2004_1\Coupon\od_c_2004_1.sav'.
MATCH FILES /FILE=*
 /RENAME (Coupons Dest Distance Origin = d0 d1 d2 d3)
 /FILE='AssignedDir\2004\2004_1\Market\od_m_2004_1.sav'
 /RENAME (ItinID Quarter Year = d4 d5 d6)
 /BY MktID
 /DROP= d0 d1 d2 d3 d4 d5 d6.
EXECUTE.
*-----------------------------------------------------------------------------------------------------------------------.
```

```
* Delete unwanted fields
*------------------------------------------------------------------------------------------------------------------------.
FILTER OFF.
USE ALL.
SELECT IF(FareClass ~= ' ' & MktFare  > 0).
EXECUTE .
*------------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Itn ID in an ascending order
*------------------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*------------------------------------------------------------------------------------------------------------------------.

SAVE OUTFILE='AssignedDir\2004\2004_1\od_cm_2004_1.sav'
 /COMPRESSED.
*------------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON MARKET & TICKET USING Itn ID
*------------------------------------------------------------------------------------------------------------------------.
GET
 FILE='AssignedDir\2004\2004_1\od_cm_2004_1.sav'.
MATCH FILES /FILE=*
 /RENAME (CouponType MktDistance MktFare MktID Origin Passengers Quarter RPCarrier SeqNum Year = d0 d1 d2 d3 d4 d5 d6 d7
d8
  d9)
 /FILE='D:\DB1B\2004\2004_1\Ticket\od_t_2004_1.sav'
 /RENAME (Coupons = d10)
 /BY ItinID
 /DROP= d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10.
```

```
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 0).
EXECUTE .

SAVE OUTFILE='AssignedDir\2004\2004_1\od_cmt_2004_1.sav'
 /COMPRESSED.
 *-----------------------------------------------------------------------------------------------------------------------.

 *-----------------------------------------------------------------------------------------------------------------------.
* DATABASE = TICKET YEAR = 2004   QUARETER = 2.
*-----------------------------------------------------------------------------------------------------------------------.
*  Import Ticket data for Quarter 1, Delete unwanted fields & Save file as a 'od_t_XXXX_2.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_2\Ticket\Origin_and_Destination_Survey_DB1BTicket_2004_2.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 Coupons F1.0
 Year F4.2
```

```
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A10
 OriginWac F2.1
 RoundTrip F4.2
 OnLine F4.2
 DollarCred F1.0
 ItinYield F6.2
 RPCarrier A2
 Passengers F5.2
 ItinFare F7.2
 BulkFare F4.2
 Distance F7.2
 DistanceGroup F2.1
 MilesFlown F7.2
 ItinGeoType F1.0
 V25 F1.0
 .
EXECUTE.

*-----------------------------------------------------------------------------------------------------------------------.
*  Delete unwanted variables.
*-----------------------------------------------------------------------------------------------------------------------.
```

```
DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac OnLine DollarCred BulkFare DistanceGroup MilesFlown ItinGeoType V25 .

*-----------------------------------------------------------------------------------------------------------------.
* Sort Cases by It ID.
*-----------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*-----------------------------------------------------------------------------------------------------------------.
* Delete all fares less than $ 50 & greater than $ 5000
*-----------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 50 & ItinFare <= 5000).
EXECUTE .

*-----------------------------------------------------------------------------------------------------------------.
* Restrict Distance from 100 miles to  10000 miles
*-----------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(Distance >= 100 & Distance <= 10000).
EXECUTE .
*-----------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_t_XXXX_2.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------.
```

```
SAVE OUTFILE='AssignedDir\2004\2004_2\Ticket\od_t_2004_2.sav'
 /COMPRESSED.
*------------------------------------------------------------------------------------------------------------------------.

* DATABASE = Market YEAR = 2004   QUARETER = 2.
*----------------------------------------------------------------------------------------------------------------.
*  Import Market data for Quarter 1, Delete unwanted fields & Save file as a 'od_m_XXXX_2.sav' file where XXXX = year.
*----------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_2\Market\Origin_and_Destination_Survey_DB1BMarket_2004_2.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 MktID F12.0
 MktCoupons F1.0
 Year F4.2
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
```

OriginStateName A14

OriginWac F2.1

Dest A3

DestAptInd F1.0

DestCityNum F5.2

DestCountry A2

DestStateFips F2.1

DestState A2

DestStateName A20

DestWac F2.1

AirportGroup A19

WacGroup A14

TkCarrierChange F4.2

TkCarrierGroup A11

OpCarrierChange F4.2

OpCarrierGroup A11

RPCarrier A2

TkCarrier A2

OpCarrier F2.1

BulkFare F4.2

Passengers F4.2

MktFare F7.2

MktDistance F7.2

MktDistanceGroup F2.1

MktMilesFlown F7.2

NonStopMiles F7.2

ItinGeoType F1.0

MktGeoType F1.0

V40 F1.0

```
 .
EXECUTE.

*-------------------------------------------------------------------------------------------------------------------.
*  Delete insignificant variables
*-------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES MktCoupons OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac  WacGroup TkCarrier  MktDistanceGroup
 TkCarrierChange  TkCarrierGroup  OpCarrierChange  OpCarrierGroup  OpCarrier Passengers ItinGeoType BulkFare MktGeoType V40 MktMilesFlown
  NonStopMiles.

*-------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*-------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*-------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_m_XXXX_2.sav' file where XXXX = year.
*-------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_2\Market\od_m_2004_2.sav'
 /COMPRESSED.
*---------------------------------------------------------------------------------------------------------------------------.
* DATABASE = Coupon YEAR = 2004   QUARETER = 2.
*-------------------------------------------------------------------------------------------------------------------.
*  Import Coupon data for Quarter 1, Delete unwanted fields & Save file as a 'od_c_XXXX_2.sav' file where XXXX = year.
*-------------------------------------------------------------------------------------------------------------------.
```

```
GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_2\Coupon\Origin_and_Destination_Survey_DB1BCoupon_2004_2.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = ""
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 MktID F12.0
 SeqNum F1.0
 Coupons F1.0
 Year F4.0
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A12
 OriginWac F2.1
 Dest A3
 DestAptInd F1.0
 DestCityNum F5.2
 DestCountry A2
 DestStateFips F2.1
```

```
 DestState A2
 DestStateName A12
 DestWac F2.1
 Break A1
 CouponType A1
 TkCarrier A2
 OpCarrier A2
 RPCarrier A2
 Passengers F4.0
 FareClass A1
 Distance F7.2
 DistanceGroup F1.0
 Gateway F4.2
 ItinGeoType F1.0
 CouponGeoType F1.0
 V35 F1.0
 .
EXECUTE.

*-----------------------------------------------------------------------------------------------------------------------.
*  DELETE UNWANTED VARIABLES
*-----------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac
 Break TkCarrier OpCarrier RPCarrier DistanceGroup Gateway ItinGeoType CouponGeoType V35.
*-----------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*-----------------------------------------------------------------------------------------------------------------------.
```

```
sort cases by MktID (a) .

*----------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_c_XXXX_2.sav' file where XXXX = year.
*----------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_2\Coupon\od_c_2004_2.sav'
 /COMPRESSED.


*----------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON & MARKET USING MARKET ID
*----------------------------------------------------------------------------------------------------------------------.

GET
 FILE='AssignedDir\2004\2004_2\Coupon\od_c_2004_2.sav'.
MATCH FILES /FILE=*
 /RENAME (Coupons Dest Distance Origin = d0 d1 d2 d3)
 /FILE='AssignedDir\2004\2004_2\Market\od_m_2004_2.sav'
 /RENAME (ItinID Quarter Year = d4 d5 d6)
 /BY MktID
 /DROP= d0 d1 d2 d3 d4 d5 d6.
EXECUTE.
*----------------------------------------------------------------------------------------------------------------------.
* Delete unwanted fields
*----------------------------------------------------------------------------------------------------------------------.
FILTER OFF.
USE ALL.
SELECT IF(FareClass ~= ' ' & MktFare  > 0).
```

```
EXECUTE .
*-------------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Itn ID in an ascending order
*-------------------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*-------------------------------------------------------------------------------------------------------------------------.

SAVE OUTFILE='AssignedDir\2004\2004_2\od_cm_2004_2.sav'
 /COMPRESSED.
*-------------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON MARKET & TICKET USING Itn ID
*-------------------------------------------------------------------------------------------------------------------------.
GET
 FILE='AssignedDir\2004\2004_2\od_cm_2004_2.sav'.
MATCH FILES /FILE=*
 /RENAME (CouponType MktDistance MktFare MktID Origin Passengers Quarter RPCarrier SeqNum Year = d0 d1 d2 d3 d4 d5 d6 d7
d8
  d9)
 /FILE='D:\DB1B\2004\2004_2\Ticket\od_t_2004_2.sav'
 /RENAME (Coupons = d10)
 /BY ItinID
 /DROP= d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 0).
```

```
EXECUTE .

SAVE OUTFILE='AssignedDir\2004\2004_2\od_cmt_2004_2.sav'
 /COMPRESSED.
 *---------------------------------------------------------------------------------------------------------------------.

 *---------------------------------------------------------------------------------------------------------------------.
* DATABASE = TICKET YEAR = 2004   QUARETER = 3.
*---------------------------------------------------------------------------------------------------------------------.
*  Import Ticket data for Quarter 1, Delete unwanted fields & Save file as a 'od_t_XXXX_3.sav' file where XXXX = year.
*---------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_3\Ticket\Origin_and_Destination_Survey_DB1BTicket_2004_3.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 Coupons F1.0
 Year F4.2
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
```

```
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A10
 OriginWac F2.1
 RoundTrip F4.2
 OnLine F4.2
 DollarCred F1.0
 ItinYield F6.2
 RPCarrier A2
 Passengers F5.2
 ItinFare F7.2
 BulkFare F4.2
 Distance F7.2
 DistanceGroup F2.1
 MilesFlown F7.2
 ItinGeoType F1.0
 V25 F1.0
 .
EXECUTE.

*--------------------------------------------------------------------------------------------------------------------------.
*  Delete unwanted variables.
*--------------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac OnLine DollarCred BulkFare DistanceGroup MilesFlown ItinGeoType V25 .

*--------------------------------------------------------------------------------------------------------------------------.
* Sort Cases by It ID.
```

```
*-----------------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*-----------------------------------------------------------------------------------------------------------------------.
* Delete all fares less than $ 50 & greater than $ 5000
*-----------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 50 & ItinFare <= 5000).
EXECUTE .

*-----------------------------------------------------------------------------------------------------------------------.
* Restrict Distance from 100 miles to  10000 miles
*-----------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(Distance >= 100 & Distance <= 10000).
EXECUTE .
*-----------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_t_XXXX_3.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_3\Ticket\od_t_2004_3.sav'
 /COMPRESSED.
*-------------------------------------------------------------------------------------------------------------------------------.

* DATABASE = Market YEAR = 2004   QUARETER = 3.
```

```
*-------------------------------------------------------------------------------------------------------------------.
*  Import Market data for Quarter 1, Delete unwanted fields & Save file as a 'od_m_XXXX_3.sav' file where XXXX = year.
*-------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_3\Market\Origin_and_Destination_Survey_DB1BMarket_2004_3.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 MktID F12.0
 MktCoupons F1.0
 Year F4.2
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A14
 OriginWac F2.1
 Dest A3
 DestAptInd F1.0
 DestCityNum F5.2
```

```
 DestCountry A2
 DestStateFips F2.1
 DestState A2
 DestStateName A20
 DestWac F2.1
 AirportGroup A19
 WacGroup A14
 TkCarrierChange F4.2
 TkCarrierGroup A11
 OpCarrierChange F4.2
 OpCarrierGroup A11
 RPCarrier A2
 TkCarrier A2
 OpCarrier F2.1
 BulkFare F4.2
 Passengers F4.2
 MktFare F7.2
 MktDistance F7.2
 MktDistanceGroup F2.1
 MktMilesFlown F7.2
 NonStopMiles F7.2
 ItinGeoType F1.0
 MktGeoType F1.0
 V40 F1.0
 .
EXECUTE.

*----------------------------------------------------------------------------------------------------------------------.
*  Delete insignificant variables
```

```
*-----------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES MktCoupons OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac  WacGroup TkCarrier  MktDis-
tanceGroup
 TkCarrierChange  TkCarrierGroup  OpCarrierChange  OpCarrierGroup  OpCarrier Passengers ItinGeoType BulkFare MktGeoType
V40 MktMilesFlown
  NonStopMiles.

*-----------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*-----------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*-----------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_m_XXXX_3.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_3\Market\od_m_2004_3.sav'
 /COMPRESSED.
*------------------------------------------------------------------------------------------------------------------------.
* DATABASE = Coupon YEAR = 2004   QUARETER = 3.
*-----------------------------------------------------------------------------------------------------------------.
*  Import Coupon data for Quarter 1, Delete unwanted fields & Save file as a 'od_c_XXXX_3.sav' file where XXXX = year.
*-----------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_3\Coupon\Origin_and_Destination_Survey_DB1BCoupon_2004_3.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
```

```
/QUALIFIER = ""
/ARRANGEMENT = DELIMITED
/FIRSTCASE = 2
/IMPORTCASE = ALL
/VARIABLES =
ItinID F12.0
MktID F12.0
SeqNum F1.0
Coupons F1.0
Year F4.0
Quarter F1.0
Origin A3
OriginAptInd F1.0
OriginCityNum F5.2
OriginCountry A2
OriginStateFips F2.1
OriginState A2
OriginStateName A12
OriginWac F2.1
Dest A3
DestAptInd F1.0
DestCityNum F5.2
DestCountry A2
DestStateFips F2.1
DestState A2
DestStateName A12
DestWac F2.1
Break A1
CouponType A1
```

```
 TkCarrier A2
 OpCarrier A2
 RPCarrier A2
 Passengers F4.0
 FareClass A1
 Distance F7.2
 DistanceGroup F1.0
 Gateway F4.2
 ItinGeoType F1.0
 CouponGeoType F1.0
 V35 F1.0
 .
EXECUTE.

*--------------------------------------------------------------------------------------------------------------------.
*  DELETE UNWANTED VARIABLES
*--------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac
 Break TkCarrier OpCarrier RPCarrier DistanceGroup Gateway ItinGeoType CouponGeoType V35.
*--------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*--------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*--------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_c_XXXX_3.sav' file where XXXX = year.
```

```
*-----------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_3\Coupon\od_c_2004_3.sav'
 /COMPRESSED.



*-----------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON & MARKET USING MARKET ID
*-----------------------------------------------------------------------------------------------------------------------.

GET
 FILE='AssignedDir\2004\2004_3\Coupon\od_c_2004_3.sav'.
MATCH FILES /FILE=*
 /RENAME (Coupons Dest Distance Origin = d0 d1 d2 d3)
 /FILE='AssignedDir\2004\2004_3\Market\od_m_2004_3.sav'
 /RENAME (ItinID Quarter Year = d4 d5 d6)
 /BY MktID
 /DROP= d0 d1 d2 d3 d4 d5 d6.
EXECUTE.
*-----------------------------------------------------------------------------------------------------------------------.
* Delete unwanted fields
*-----------------------------------------------------------------------------------------------------------------------.
FILTER OFF.
USE ALL.
SELECT IF(FareClass ~= ' ' & MktFare  > 0).
EXECUTE .
*-----------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Itn ID in an ascending order
*-----------------------------------------------------------------------------------------------------------------------.
```

```
sort cases by ItinID (a) .

*-----------------------------------------------------------------------------------------------------------------------.

SAVE OUTFILE='AssignedDir\2004\2004_3\od_cm_2004_3.sav'
 /COMPRESSED.
*-----------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON MARKET & TICKET USING Itn ID
*-----------------------------------------------------------------------------------------------------------------------.
GET
 FILE='AssignedDir\2004\2004_3\od_cm_2004_3.sav'.
MATCH FILES /FILE=*
 /RENAME (CouponType MktDistance MktFare MktID Origin Passengers Quarter RPCarrier SeqNum Year = d0 d1 d2 d3 d4 d5 d6 d7
d8
  d9)
 /FILE='D:\DB1B\2004\2004_3\Ticket\od_t_2004_3.sav'
 /RENAME (Coupons = d10)
 /BY ItinID
 /DROP= d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 0).
EXECUTE .

SAVE OUTFILE='AssignedDir\2004\2004_3\od_cmt_2004_3.sav'
 /COMPRESSED.
 *-----------------------------------------------------------------------------------------------------------------------.
```

```
*----------------------------------------------------------------------------------------------------------------------.
* DATABASE = TICKET YEAR = 2004   QUARETER = 4.
*----------------------------------------------------------------------------------------------------------------------.
*  Import Ticket data for Quarter 1, Delete unwanted fields & Save file as a 'od_t_XXXX_4.sav' file where XXXX = year.
*----------------------------------------------------------------------------------------------------------------------.
GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_4\Ticket\Origin_and_Destination_Survey_DB1BTicket_2004_4.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = ''''
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
 Coupons F1.0
 Year F4.2
 Quarter F1.0
 Origin A3
 OriginAptInd F1.0
 OriginCityNum F5.2
 OriginCountry A2
 OriginStateFips F2.1
 OriginState A2
 OriginStateName A10
 OriginWac F2.1
 RoundTrip F4.2
 OnLine F4.2
```

```
 DollarCred F1.0
 ItinYield F6.2
 RPCarrier A2
 Passengers F5.2
 ItinFare F7.2
 BulkFare F4.2
 Distance F7.2
 DistanceGroup F2.1
 MilesFlown F7.2
 ItinGeoType F1.0
 V25 F1.0
 .
EXECUTE.

*-----------------------------------------------------------------------------------------------------------------------.
*  Delete unwanted variables.
*-----------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac OnLine DollarCred BulkFare DistanceGroup MilesFlown ItinGeoType V25 .

*-----------------------------------------------------------------------------------------------------------------------.
* Sort Cases by It ID.
*-----------------------------------------------------------------------------------------------------------------------.

sort cases by ItinID (a) .

*-----------------------------------------------------------------------------------------------------------------------.
* Delete all fares less than $ 50 & greater than $ 5000
```

```
*-------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 50 & ItinFare <= 5000).
EXECUTE .

*-------------------------------------------------------------------------------------------------------------------.
* Restrict Distance from 100 miles to  10000 miles
*-------------------------------------------------------------------------------------------------------------------.

FILTER OFF.
USE ALL.
SELECT IF(Distance >= 100 & Distance <= 10000).
EXECUTE .
*-------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_t_XXXX_4.sav' file where XXXX = year.
*-------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_4\Ticket\od_t_2004_4.sav'
 /COMPRESSED.
*---------------------------------------------------------------------------------------------------------------------------.

* DATABASE = Market YEAR = 2004   QUARETER = 4.
*-------------------------------------------------------------------------------------------------------------------.
*  Import Market data for Quarter 1, Delete unwanted fields & Save file as a 'od_m_XXXX_4.sav' file where XXXX = year.
*-------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_4\Market\Origin_and_Destination_Survey_DB1BMarket_2004_4.csv'
```

```
/DELCASE = LINE
/DELIMITERS = ","
/QUALIFIER = ""
/ARRANGEMENT = DELIMITED
/FIRSTCASE = 2
/IMPORTCASE = ALL
/VARIABLES =
ItinID F12.0
MktID F12.0
MktCoupons F1.0
Year F4.2
Quarter F1.0
Origin A3
OriginAptInd F1.0
OriginCityNum F5.2
OriginCountry A2
OriginStateFips F2.1
OriginState A2
OriginStateName A14
OriginWac F2.1
Dest A3
DestAptInd F1.0
DestCityNum F5.2
DestCountry A2
DestStateFips F2.1
DestState A2
DestStateName A20
DestWac F2.1
AirportGroup A19
```

```
 WacGroup A14
 TkCarrierChange F4.2
 TkCarrierGroup A11
 OpCarrierChange F4.2
 OpCarrierGroup A11
 RPCarrier A2
 TkCarrier A2
 OpCarrier F2.1
 BulkFare F4.2
 Passengers F4.2
 MktFare F7.2
 MktDistance F7.2
 MktDistanceGroup F2.1
 MktMilesFlown F7.2
 NonStopMiles F7.2
 ItinGeoType F1.0
 MktGeoType F1.0
 V40 F1.0
 .
EXECUTE.

*------------------------------------------------------------------------------------------------------------------------.
*  Delete insignificant variables
*------------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES MktCoupons OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac  WacGroup TkCarrier  MktDistanceGroup
 TkCarrierChange  TkCarrierGroup  OpCarrierChange  OpCarrierGroup  OpCarrier Passengers ItinGeoType BulkFare MktGeoType V40 MktMilesFlown
```

```
  NonStopMiles.

*------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_m_XXXX_4.sav' file where XXXX = year.
*------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_4\Market\od_m_2004_4.sav'
 /COMPRESSED.
*-------------------------------------------------------------------------------------------------------------------------.
* DATABASE = Coupon YEAR = 2004   QUARETER = 4.
*------------------------------------------------------------------------------------------------------------------.
*  Import Coupon data for Quarter 1, Delete unwanted fields & Save file as a 'od_c_XXXX_4.sav' file where XXXX = year.
*------------------------------------------------------------------------------------------------------------------.

GET DATA  /TYPE = TXT
 /FILE = 'AssignedDir\2004\2004_4\Coupon\Origin_and_Destination_Survey_DB1BCoupon_2004_4.csv'
 /DELCASE = LINE
 /DELIMITERS = ","
 /QUALIFIER = '"'
 /ARRANGEMENT = DELIMITED
 /FIRSTCASE = 2
 /IMPORTCASE = ALL
 /VARIABLES =
 ItinID F12.0
```

MktID F12.0

SeqNum F1.0

Coupons F1.0

Year F4.0

Quarter F1.0

Origin A3

OriginAptInd F1.0

OriginCityNum F5.2

OriginCountry A2

OriginStateFips F2.1

OriginState A2

OriginStateName A12

OriginWac F2.1

Dest A3

DestAptInd F1.0

DestCityNum F5.2

DestCountry A2

DestStateFips F2.1

DestState A2

DestStateName A12

DestWac F2.1

Break A1

CouponType A1

TkCarrier A2

OpCarrier A2

RPCarrier A2

Passengers F4.0

FareClass A1

Distance F7.2

```
 DistanceGroup F1.0
 Gateway F4.2
 ItinGeoType F1.0
 CouponGeoType F1.0
 V35 F1.0
 .
EXECUTE.

*------------------------------------------------------------------------------------------------------------------.
*  DELETE UNWANTED VARIABLES
*------------------------------------------------------------------------------------------------------------------.

DELETE VARIABLES OriginAptInd OriginCityNum OriginCountry OriginStateFips OriginState OriginStateName
 OriginWac DestAptInd DestCityNum DestCountry DestStateFips DestState DestStateName DestWac
 Break TkCarrier OpCarrier RPCarrier DistanceGroup Gateway ItinGeoType CouponGeoType V35.
*------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Market ID in an ascending order
*------------------------------------------------------------------------------------------------------------------.

sort cases by MktID (a) .

*------------------------------------------------------------------------------------------------------------------.
*  Save file as a 'od_c_XXXX_4.sav' file where XXXX = year.
*------------------------------------------------------------------------------------------------------------------.
SAVE OUTFILE='AssignedDir\2004\2004_4\Coupon\od_c_2004_4.sav'
 /COMPRESSED.


*------------------------------------------------------------------------------------------------------------------.
```

```
*  COMBINE COUPON & MARKET USING MARKET ID
*-----------------------------------------------------------------------------------------------------------------------.


GET
 FILE='AssignedDir\2004\2004_4\Coupon\od_c_2004_4.sav'.
MATCH FILES /FILE=*
 /RENAME (Coupons Dest Distance Origin = d0 d1 d2 d3)
 /FILE='AssignedDir\2004\2004_4\Market\od_m_2004_4.sav'
 /RENAME (ItinID Quarter Year = d4 d5 d6)
 /BY MktID
 /DROP= d0 d1 d2 d3 d4 d5 d6.
EXECUTE.
*-----------------------------------------------------------------------------------------------------------------------.
* Delete unwanted fields
*-----------------------------------------------------------------------------------------------------------------------.
FILTER OFF.
USE ALL.
SELECT IF(FareClass ~= ' ' & MktFare  > 0).
EXECUTE .
*-----------------------------------------------------------------------------------------------------------------------.
* Sort Cases by Itn ID in an ascending order
*-----------------------------------------------------------------------------------------------------------------------.


sort cases by ItinID (a) .


*-----------------------------------------------------------------------------------------------------------------------.


SAVE OUTFILE='AssignedDir\2004\2004_4\od_cm_2004_4.sav'
 /COMPRESSED.
```

```
*-------------------------------------------------------------------------------------------------------------------------.
*  COMBINE COUPON MARKET & TICKET USING Itn ID
*-------------------------------------------------------------------------------------------------------------------------.
GET
 FILE='AssignedDir\2004\2004_4\od_cm_2004_4.sav'.
MATCH FILES /FILE=*
 /RENAME (CouponType MktDistance MktFare MktID Origin Passengers Quarter RPCarrier SeqNum Year = d0 d1 d2 d3 d4 d5 d6 d7
d8
  d9)
 /FILE='D:\DB1B\2004\2004_4\Ticket\od_t_2004_4.sav'
 /RENAME (Coupons = d10)
 /BY ItinID
 /DROP= d0 d1 d2 d3 d4 d5 d6 d7 d8 d9 d10.
EXECUTE.

FILTER OFF.
USE ALL.
SELECT IF(ItinFare >= 0).
EXECUTE .

SAVE OUTFILE='AssignedDir\2004\2004_4\od_cmt_2004_4.sav'
 /COMPRESSED.
*-------------------------------------------------------------------------------------------------------------------------.

*-------------------------------------------------------------------------------------------------------------------------.
*  COMBINE DB1B DATA FOR ALL THE FOUR QUARTERS
*-------------------------------------------------------------------------------------------------------------------------.
GET
  FILE='AssignedDir\2004\2004_1\od_cmt_2004_1.sav'.
```

```
ADD FILES /FILE=*
 /FILE='AssignedDir\2004\2004_2\od_cmt_2004_2.sav'.
EXECUTE.

SAVE OUTFILE='AssignedDir\2004\od_cmt_2004.sav'
 /COMPRESSED.
*---------------------------------------------------------------------------------------------------------------------.

GET
  FILE='AssignedDir\2004\od_cmt_2004.sav'.

ADD FILES /FILE=*
 /FILE='AssignedDir\2004\2004_3\od_cmt_2004_3.sav'.
EXECUTE.

SAVE OUTFILE='AssignedDir\2004\od_cmt_2004.sav'
 /COMPRESSED.
*---------------------------------------------------------------------------------------------------------------------.
GET
  FILE='AssignedDir\2004\od_cmt_2004.sav'.
ADD FILES /FILE=*
 /FILE='AssignedDir\2004\2004_4\od_cmt_2004_4.sav'.
EXECUTE.

SAVE OUTFILE='AssignedDir\2004\od_cmt_2004.sav'
 /COMPRESSED.
*---------------------------------------------------------------------------------------------------------------------.
```

```
*--------------------------READ ME--------------------------------.
* This syntax was originally written by Gwilym Price of University of
* Glasgow. This syntax is used to test and correct heteroscedasticity.
* This syntax was updated by Krishna Murthy at Virginia Tech for SATS lab.
*--------------------------READ ME--------------------------------.

REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT lnfare
  /METHOD=ENTER lndist_c lnpax lnhin_ori lnhin_des dummy_lc ori_type
  dest_type
  /SAVE RESID(RES_1) .
*-------------------------------------------------------------------.
COMPUTE ESQ = RES_1 * RES_1.
EXECUTE.

*-------------------------------------------------------------------.
FILTER OFF.
USE ALL.
EXECUTE .
COMPUTE CONSTANT = 1.
EXECUTE.

*-------------------------------------------------------------------.
FILTER OFF.
USE ALL.
SELECT IF(MISSING(ESQ) = 0).
EXECUTE .
```

```
*--------------------------------------------------------------------.

MATRIX.
GET Y / VARIABLES = lnfare.
GET X / VARIABLES = CONSTANT, lndist_c, lnpax, lnhin_ori, lnhin_des, dummy_lc, ori_type
/ NAMES = XTITLES.
GET RESIDUAL / VARIABLES = RES_1.
GET ESQ / VARIABLES = ESQ.
COMPUTE XRTITLES = TRANSPOS(XTITLES).
COMPUTE N = NROW(ESQ).
COMPUTE K = NCOL(X).
COMPUTE O = MDIAG(ESQ).
COMPUTE WHITEV = (INV(TRANSPOS(X) * X)) *TRANSPOS(X)* O *
X*INV(TRANSPOS(X) * X).
COMPUTE WDIAG = DIAG(WHITEV).
COMPUTE WHITE_SE = SQRT(WDIAG).
PRINT WHITE_SE
/ FORMAT = "E13"
/ TITLE = "White's (Large Sample) Corrected Standard Errors"
/ RNAMES = XRTITLES.
COMPUTE B = (INV(TRANSPOS(X) * X)) * (TRANSPOS(X) * Y).
PRINT B
/ FORMAT = "E13"
/TITLE = "OLS Coefficients"
/ RNAMES = XRTITLES.
COMPUTE WT_VAL = B / WHITE_SE.
PRINT WT_VAL
/ FORMAT = "E13"
/ TITLE = "t-values based on Whites (large sample) corrected SEs"
/ RNAMES = XRTITLES.
```

```
COMPUTE SIG_WT = 2*(1- TCDF(ABS(WT_VAL), N)) .
PRINT SIG_WT
/ FORMAT = "E13"
/ TITLE = "Prob(t < tc) based on Whites (large n) SEs"
/ RNAMES = XRTITLES.
COMPUTE SIGMASQ = (TRANSPOS(RESIDUAL)*RESIDUAL)/(N-K).
COMPUTE SE_SQ = SIGMASQ*INV(TRANSPOS(X)*X).
COMPUTE SESQ_ABS = ABS(SE_SQ).
COMPUTE SE = SQRT(DIAG(SESQ_ABS)).
PRINT SE
/ FORMAT = "E13"
/ TITLE = "OLS Standard Errors"
/ RNAMES = XRTITLES.
COMPUTE OLST_VAL = B / SE.
PRINT OLST_VAL
/ FORMAT = "E13"
/ TITLE = "OLS t-values"
/ RNAMES = XRTITLES.
COMPUTE SIG_OLST = 2*(1- TCDF(ABS(OLST_VAL), N)) .
PRINT SIG_OLST
/ FORMAT = "E13"
/ TITLE = "Prob(t < tc) based on OLS SEs"
/ RNAMES = XRTITLES.
COMPUTE WESTIM = {B, SE, WHITE_SE, WT_VAL, SIG_WT}.
PRINT WESTIM
/ FORMAT = "E13"
/ RNAMES = XRTITLES
/ CLABELS = B, SE, WHITE_SE, WT_VAL, SIG_WT.
END MATRIX.
```

Bibliography

1)William P. Anderson, Gang Gong and T.R. Lakshmanan. Geographical Variation in Cost of Air travel. Transportation Research Record 1788.

2) Gang Gong. Analysis of US Domestic Air Travel Cost using GIS and Spatial Analysis. ESRI Transportation papers.

3)Gautam Gowrisankaran. Competition and Regulation in the Airline Industry. FRBSF Economic letter # 2002-01, January 18, 2002.

4)Morrison, Steven A & Winston, Clifford, 1990. The Dynamics of Airline Pricing and Competition. American Economic Review, Vol. 80 (2) pp. 389-93.

5) Charles Najda. Low-Cost Carriers and Low Fares: Competition and Concentration in the U.S. Airline Industry.

6)Sheng-Chen Alex Huang, 2000. An Analysis of Air Passenger Average Trip Lengths and Fare Levels in US Domestic Markets. ITS February 2000.

7) Gwilym Price. Heteroscedasticity: Testing and Correcting in SPSS. www.spsstools.net.

8) ________, Bureau of Transportation Statistics, “Airline Origin and Destination Survey”, http://transtats.bts.gov, (2000).

9) _______, Bureau of Transportation Statistics, “Official Airline Guide”, http://transtats.bts.gov, (2000).

10) U.S. Department of Transportation, Dominated Hub fare.

11)Thorsten Fischer and Davis R. Kamerschen. Measuring Competition in the U.S. Airline industry using the Rosse-Panzar test and Cross-sectional Regression analyses. Journal of Applied Economics Vol. VI No. 1 (August 2002), 73-93.

12)Small Aircraft Transportation Systems (SATS) (www.sats.nasa.gov).

13)Introduction to Linear Regression Analysis by Montgomery, Peck & Vining.

14) Chris Wild and George Seber. Chap. 10 of CHANCE ENCOUNTERS: A First Course in Data Analysis and Inference.