

## Airline Revenue Optimization Problem: a Multiple Linear Regression Model

Betty René Ferguson  
Social Security Administration  
6401 Security Blvd.  
Baltimore, MD 21235, USA  
and  
Don Hong  
Department of Mathematical Sciences  
Middle Tennessee State University  
Murfreesboro, TN 37132, USA  
dhong@mtsu.edu

### Abstract

Airline yield management has been a topic of research since the deregulation of the airline industry in the 1970's. The goal of airline yield management is to optimize seat allocations of a flight among the different fare products. In this paper, we use econometrics modeling to construct market demand functions. Then multiple linear regression is applied to the market demand functions. The use of multiple linear regression allows for an improved discussion of elasticity, cost degradation, and passenger diversion. A model is then constructed to optimize revenue for domestic flights. This paper answers the following specific research questions: How does the use of price elasticities and cross price elasticities improve previous models? Does the use of income elasticity improve the market demand functions? Conditions for optimality are then discussed using the estimated market demand functions.

*Key words:* airline revenue, cost degradation, econometrics modeling, elasticity, multiple linear regression, optimization.

**2000 MSC Codes:** 78M50, 62J12.

## 1 Introduction

Airline yield management, a hot topic of research since the 1970's, is used to optimize seat allocations of a single flight among the different fare products. Most models for airline yield management can be grouped into one of the following two categories: a price discrimination model or a product differentiation model. Price discrimination models assume that when a consumer chooses to purchase a lower priced fare product they do so at no additional cost. If the lower priced fare product requires a purchase of 14 days in advance or any other restrictions applied to a discount purchase, which would not have been encountered by a higher priced fare product, the assumption states that there is no cost to the consumer for accepting more restrictions.

There is an extensive study by Morrison and Winston [6] which estimates the additional costs for accepting more restrictions. Their study supports the need to eliminate the assumption imposed by price discrimination models. The other category, product differentiation models, assumes the demand for fare product  $i$  is independent of the demand for fare product  $j$  and independent of the price of any other fare products. This paper supports the need to eliminate the assumptions imposed by both the price and product discrimination models.

Botimer and Belobaba [2] introduced a generalized cost model of airline fare product differentiation. Their model for air travel demand in an origin-destination market is extended to include degradation costs and passenger diversion. These extensions eliminate the unrealistic assumptions made by previous price discrimination models and product differentiation models. Degradation costs are the costs to consumers who wish to downgrade to a lower fare product. The lower fare product requires the acceptance of a restriction(s) which may come as a cost to the consumer. Their initial demand function without degradation costs is

$$Q_i = f_i(P_i) - \sum_{j=1}^{i-1} Q_j,$$

where  $Q_i$  denotes the number of passengers purchasing fare product  $i$ ,  $f_i(\cdot)$  is the market demand function for fare product  $i$ ,  $P_i$  is the price of fare product  $i$ , and  $Q_j$  is the number of passengers held captive to fare products less restricted than fare product  $i$ .

Note that fare product  $i + 1$  is defined to impose more restrictions than fare product  $i$ . The model in [2] is designed with the following criteria:

- (1)  $f_{i+1} < f_i < f_j$  for  $j < i$ .
- (2) The market demand function is a positive function determined by customers, competitors, prices, etc. and exploited by the individual airlines.
- (3) The consumers arrive in increasing order of willingness to pay.
- (4) Demand for fare product  $i$  is derived from unrestricted fare product 1.

To include degradation costs, costs associated with consumers for accepting

more restrictions, their demand function then becomes

$$Q_{i+1} = f_{i+1}(P_{i+1} + \sum_{j=1}^{i-1} c_j) - \sum_{j=1}^{i-1} Q_j,$$

where,  $c_i$  is the cost to each consumer for accepting the imposed restrictions. Note that  $c_1$  is zero because there are no restrictions for the full fare product, and thus no cost to consumers. Their model assumes  $c_i$  is a constant cost functional form for simplicity reasons. The consumers perceived cost of fare product  $i$  is higher than the actual price of fare product  $i$ . Thus as Botimer and Belobaba state in [2], “their willingness to purchase fare product  $i$  is reduced by  $c_i$  as compared to fare product  $i - 1$ . ”

The model in [2] for air travel demand is extended to include passenger diversion to eliminate the assumption of previous product differentiation models. This is designed to include a fixed percentage of the expected demand for any fare product. Thus the number of passengers actually purchasing fare product  $i$  is represented by  $q_i$ ,

$$q_i = (1 - \sum_{j=i+1}^N d_{ij})Q_i + \sum_{j=1}^{i-1} d_{ji}Q_j,$$

where,  $d_{ij}$  denotes the percentage of passengers diverting from fare product  $i$  to a more restricted fare product  $j$ .

Finally, their optimizing revenue function is constructed only for the linear case of the constant cost model where,

$$P_i = P_0 - a \sum_{j=1}^{i-1} Q_j - \sum_{j=1}^{i-1} c_j$$

is strictly nonincreasing. The Lagrange Multiplier Method is used to maximize the following revenue objective function which includes degradation costs and passenger diversion:

$$R = \sum_{i=1}^N (1 - \sum_{i=1}^N d_{ij}) Q_{ij} [P_0 - a \sum_{k=1}^i Q_k - \sum_{r=1}^i c_r] + \sum_{i=1}^N \sum_{j=i+1}^N d_{ji} Q_j [P_0 - a \sum_{k=1}^j Q_k - \sum_{r=1}^j c_r].$$

This model leaves room for improvement as most models do. In [2], the authors mentioned several areas for further research. Under the topic on passenger diversion, the authors suggested the use of cross price elasticity effects in a model. Under the topic on degradation costs, the authors suggested that constant cost formulation does not realistically reflect consumer behavior. That is, “costs incurred may differ by passenger rather than being constant.” So there is

a need to determine the effects of passenger behavior. In this paper, comparing to the model in [2] constructed only for the linear case of the constant cost model, we construct a model of market demand function using econometrics modeling that could be used for forecasting consumer behavior in the airline industry. Based on a stratified random sample from the U.S. Department of Transportation's Domestic Airline Fares Consumer Report of 1997, we determine the market demand functions including price and cross price elasticity with and without income elasticity using multiple linear regression. We further analyze these results and determine a generalized objective function. In the final section, we apply Lagrange multiplier method to solve the the airline revenue maximization problems.

## 2 Market Demand Function

Econometrics modeling of the air travel demand will allow us to observe cost degradation and passenger diversion in action. Instead of fixing the percentage of diversions between the fare products and having a constant degradation cost  $c_i$ , we shall model existing behaviors in hopes to be able to better forecast future consumer behavior. For research of econometric modeling, we refer to the book [7]: *Econometric Models and Econometric Forecasts*.

To understand econometrics modeling and how this can work for airtravel demand, we first recall some basic concepts of elasticity. An elasticity measures the effect on the dependent variable of a 1 percent change in an independent variable. Therefore, we can monitor change of the dependent variable  $Q_i$  of a 1 percent change in an independent variable  $P_i$ , where  $Q_i$  is demand for a product and  $P_i$  is the price for this product. This situation is called price elasticity. We can also monitor  $Q_i$  of a 1 percent change in another independent variable  $P_j$ . This situation is called cross price elasticity. Elasticities are easy to work with due to the facts that their values are unbounded, values may be positive or negative, and are unit-free. A market demand function which includes price elasticity and cross price elasticity may prove to be a more realistic approach to consumer behavior.

Econometric modeling for demand yields the following equation for fare product  $i$

$$Q_i = \beta_{i0} P_i^{\beta_i} \prod_{i \neq j} P_j^{\beta_{ij}} e^{\epsilon_i},$$

where,  $Q_i$  is a continuous, dependent variable representing quantity demanded for fare product  $i$ ,  $\beta_{i0}$  is an unbounded and unit-free constant,  $\beta_i$  is the price elasticity which is unbounded and unit-free for fare product  $i$ ,  $\beta_{ij}$  is the unbounded and unit-free cross price elasticity for fare product  $i$  by change in price  $j$ ,  $P_i$  is an independent variable representing price of fare product  $i$ , and  $\epsilon_i$  is the error term which assumes a normal distribution.

An econometric model of airline demand shall yield as many equations as there are fare products. Thus for simplicity purposes we shall model demand

aggregating passenger services into three fare products. That is, fare product 1 will have demand function  $Q_1$  representing demand for first class or full fare products, fare product 2 will have demand function  $Q_2$  representing demand for standard economy class or first class with some restrictions, and fare product 3 will have demand function  $Q_3$  representing the demand for discount fareclass or standard economy class with many more restrictions.  $Q_2$  usually requires advance purchase of 3 days while  $Q_3$  usually requires an advance purchase of 14 or more days.

It is clear that we only want to include necessary independent variables. It is not necessary to include  $P_3$  in  $Q_1$  due to the fact that demand in first class is not effected by price changes of fares in the discount fareclass. However, as we will see, we must include  $P_1$ ,  $P_2$ , and  $P_3$  in  $Q_2$  and only  $P_3$  and  $P_2$  in  $Q_3$ . Realistically, demand should be effected by the above fareclass and below fareclass changes in price. Thus standard economy fareclass, in the three fare product model, is the only fare product that has two cross price elasticities. In order to linearize the model, we take the natural log of each  $Q_i$  yielding the following three fare product market demand functions:

$$\begin{aligned}\ln Q_1 &= \beta_{10} + \beta_1 \ln P_1 + \beta_{12} \ln P_2 + \epsilon_1, \\ \ln Q_2 &= \beta_{20} + \beta_2 \ln P_2 + \beta_{21} \ln P_1 + \beta_{23} \ln P_3 + \epsilon_2, \\ \ln Q_3 &= \beta_{30} + \beta_3 \ln P_3 + \beta_{32} \ln P_2 + \epsilon_3.\end{aligned}$$

For the multiple linear regression model, we let  $q_i = \ln Q_i$  and  $x_i = \ln P_i$ . Thus we have the following three market demand functions:

$$\begin{aligned}q_1 &= \beta_{10} + \beta_1 x_1 + \beta_{12} x_2 + \epsilon_1, \\ q_2 &= \beta_{20} + \beta_2 x_2 + \beta_{21} x_1 + \beta_{23} x_3 + \epsilon_2, \\ q_3 &= \beta_{30} + \beta_3 x_3 + \beta_{32} x_2 + \epsilon_3.\end{aligned}$$

The assumptions of these multiple linear regressions in [7] are: the relationship between  $q_i$  and  $x_i$  is linear, the  $x_i$ 's are non-stochastic variables and in addition, no exact linear relationship exists between two or more independent variables, the error term has zero expected value for all observations, the error term as constant variance for all observations, and the error term is normally distributed.

With these assumptions, we would like to use the model to analyze consumer behavior through the use of price elasticities and cross price elasticities. Therefore, we select a sample of approximately 250 flights to monitor consumer behavior. The data from the 250 flights are the data for  $q_i$ 's and  $x_i$ 's in the above model. We use multiple linear regression to search for parameter estimates,  $\beta$ 's, that minimize the error sums of squares.

The squared sum of errors is  $SSE_i = \sum_j (e_i^{(j)})^2 = \sum_j (q_i^{(j)} - \hat{q}_i^{(j)})^2$ , where  $q_i^{(j)}$  is the observed value for the natural log of demand of the flight  $j$  and  $\hat{q}_i^{(j)}$

is the predicted value for the natural log of the demand of the flight  $j$ . Thus we will have 250 equations for quantity demanded for each of the  $q_i$ 's with the unknown elasticities  $\beta$ 's. Using multiple linear regression we will have one predicted demand function for each of the  $q_i$ 's:

$$\hat{q}_i = \hat{\beta}_{i0} + \hat{\beta}_i x_i + \hat{\beta}_{i(i-1)} x_{i-1} + \hat{\beta}_{i(i+1)} x_{i+1}, \quad i = 1, 2, 3.$$

This model needs more than three observations, that is, three or more flights. Multiple linear regression is used to solve for  $\hat{\beta}_i$  and  $\hat{\beta}_{ij}$  for  $i \neq j$ . Then  $\hat{\beta}_{i0}$  can be solved. The parameter estimates are defined as the following for  $\hat{q}_1$  and similarly for  $\hat{q}_2$  and  $\hat{q}_3$ :

$$\hat{\beta}_1 = \frac{(\sum_{i=1}^{250} x_1^{(i)} q_1^{(i)})(\sum_{i=1}^{250} (x_2^{(i)})^2) - (\sum_{i=1}^{250} x_2^{(i)} q_1^{(i)})(\sum_{i=1}^{250} x_1^{(i)} x_2^{(i)})}{(\sum_{i=1}^{250} (x_1^{(i)})^2)(\sum_{i=1}^{250} (x_2^{(i)})^2) - (\sum_{i=1}^{250} x_1^{(i)} x_2^{(i)})^2},$$

$$\hat{\beta}_{12} = \frac{(\sum_{i=1}^{250} x_2^{(i)} q_1^{(i)})(\sum_{i=1}^{250} (x_2^{(i)})^2) - (\sum_{i=1}^{250} x_1^{(i)} q_1^{(i)})(\sum_{i=1}^{250} x_1^{(i)} x_2^{(i)})}{(\sum_{i=1}^{250} (x_1^{(i)})^2)(\sum_{i=1}^{250} (x_2^{(i)})^2) - (\sum_{i=1}^{250} x_1^{(i)} x_2^{(i)})^2},$$

and

$$\hat{\beta}_{10} = \bar{q}_1 - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_{12} \bar{x}_2,$$

where  $\bar{q}_1 = \frac{1}{250} \sum_{i=1}^{250} q_1^{(i)}$ , and  $q_1^{(i)}$  is the observed value for the natural log of demand of the flight  $i$ .  $\bar{x}_1$  and  $\bar{x}_2$  are defined similarly.

### 3 Main Results and Analysis

To estimate our price and cross price elasticities, a stratified random sample is chosen from U.S. Department of Transportation's Domestic Airline Fares Consumer Report of 1997. This report of the 1,000 largest city-pair markets within the 48 states accounts for approximately 75 percent of all 48-state passengers flights. The 1,000 flights are divided into groups determined by their nonstop distance. A separate simple random sample (SRS) is used to select from the list of flights whose nonstop distance ranges from 100–300 miles, 500–700 miles, 900–1100 miles, 1300–1500 miles, and 1900–2100 miles. The combined SRS in each category yields 250 randomly selected flights; 25 percent of the population of interest.

The prices used for the three fare class model are current prices given from various search engines comprised of [www.flyaow.com](http://www.flyaow.com), [www.travelocity.com](http://www.travelocity.com), and [www.bestlodgings.com](http://www.bestlodgings.com). These search engines allow us to determine the average prices for each of the three fare classes. The price for standard economy is determined by requiring an advanced purchase of 3 days and 14 days for discount fareclass. The model could also include average median incomes of the city of

departure and the city of arrival. Therefore, in the following, the demand functions will first be solved using price elasticity and cross price elasticity. Then the demand functions will be solved using price, cross price, and income elasticity. Finally, we use statistical analysis to determine whether income elasticity is a useful independent variable for quantity demanded.

Using SAS and applying multiple linear regression to the data for the 250 flights yields the following results:

Market Demand Functions (with price and cross price elasticity)

$$\begin{aligned}\hat{q}_1 &= 9.61 - .253x_1 - .650x_2, r^2 = .411, \\ \hat{q}_2 &= 10.8 - .550x_2 - .224x_1 - .123x_3, r^2 = .414, \\ \hat{q}_3 &= 10.7 - .172x_3 - .793x_2, r^2 = .468,\end{aligned}$$

where  $r^2$  (adj) is .406, .407, and .464 respectively.

Market Demand Functions (with price and cross price elasticity plus income elasticity)

$$\begin{aligned}\hat{q}_1 &= 5.15 - .225x_1 - .669x_2 + .423I, r^2 = .415, \\ \hat{q}_2 &= 6.38 - .570x_2 - .216x_1 - .126x_3 + .419I, r^2 = .418, \\ \hat{q}_3 &= 6.1 - .169x_3 - .791x_2 + .439I, r^2 = .473,\end{aligned}$$

where  $r^2$  (adj) is .408, .409 and .467 respectively and  $I$  is the average of the cities (departure and arrival) median family incomes.

Analysis of the model involves several different methods. The methods used here are described in [4]. First, we must analyze the assumptions of the model. One is that the random error term assumes a normal distribution. The histograms of the residuals plotted against each of the independent variables:  $q_1$ ,  $q_2$ , and  $q_3$  indicate a few outliers which may cause a lower than usual measure of fit. Overall, the three histograms appear to have a normal distribution. Thus the analysis of these histograms do not give any indication that the normality assumption of the model has not been met. The normal probability plots of the residuals against each  $q_1$ ,  $q_2$ , and  $q_3$  also show a few possible outliers. According to [4], “an outlier among residuals is one that is far greater than the rest in absolute value and perhaps lies three or four standard deviations or further from the mean of the residuals.” Thus there are some concerns from these plots that a few of the flights do not have data that are typical to the rest of the flights. Never the less, the linearity in each of the plots suggests their are no indications that the normality assumption has not been met. Also, the plots of the residuals against the fitted values for  $q_1$ ,  $q_2$ , and  $q_3$  show a few outliers. However, if we remove these outliers, our graph shows constant variance. Thus our assumptions have been met.

Second, we must analyze the cross price elasticity, to verify its importance in the model. We can analyze the analysis of variance tables (ANOVA) to

Table 1: Analysis of Variance:  $q_1$  regress on  $x_1$ 

Source	DF	Sum Sq.s	Mean Sq.	F-Value	Prob> F
Model	1	52.69197	52.69197	143.321	0.0001
Error	250	91.91282	0.36765		
C Total	251	144.60480			

  

Root MSE	0.60634	R-square	0.3644	Adj R-sq	.03618
----------	---------	----------	--------	----------	--------

  

Parameter Estimates

Variable	DF	$\beta$ Estimate	Standard Error
INTERCEP	1	10.006380	0.48941062
X1	1	-0.912193	0.07619605

  

Variable	T: $\beta = 0$	Prob >  T
INTERCEP	20.446	0.0001
X1	-11.972	0.0001

easily verify the importance of cross price elasticity. The Tables 1–3 are the ANOVA tables for  $q_1$  regressed onto each of the independent variables  $x_1$ ,  $x_2$ , and  $I$ . Most obvious from the ANOVA tables are the p-values. The p-value for  $q_1 = f(x_1)$  is nearly zero (Table 1) and the p-value for  $q_1 = f(x_2)$  is also nearly zero (Table 2). The regression of  $q_1$  on both  $x_1$  and  $x_2$  yields the following sums of squares and the proportion of variations:

$$SSR(x_1) = 48.6686, SSR(x_2|x_1) = 10.4507, r_1^2 = .3365, r_{x_2|x_1}^2 = .1089.$$

These are located in Table 4. It is clear that the demand for the full fare product relies on the price of fare product 2. The data tells us that before  $x_2$  is added to the model, the  $q_1$  with only  $x_1$  in the model had a proportion of variations of .3365. And then, once  $x_2$  is added to the model, the proportion of variations by  $x_2$  after  $x_1$  in the model becomes .1089. It suggests that the model includes cross price elasticity. The necessity for the cross price elasticity can be observed in the same way for  $q_2$  and  $q_3$ .

Thus the question still remains, what independent variable is missing from the model?  $r^2$  for all three market demand functions are in the .40 range. For observational data, there are hopes for  $r^2$  to be closer to the .60 range. Therefore, the research was expanded to check for another possible independent variable that may explain the proportion of variations in the quantity demanded. The research expanded to include income elasticity in the model. Since the consumers are purchasing a product, the income for the consumers is an obvious

Table 2: Analysis of Variance:  $q_1$  regress on  $x_2$

Source	DF	Sum Sq.s	Mean Sq.	F-Value	Prob> F
Model	1	58.61769	58.61769	170.426	0.0001
Error	250	85.98710	0.34395		
C Total	251	144.60480			

Root MSE 0.58647 R-square 0.4054 Adj R-sq 0.4030

Parameter Estimates

Variable	DF	$\beta$ Estimate	Standard Error
INTERCEP	1	9.171135	0.38523367
X2	1	-0.851072	0.06519264

  

Variable	T: $\beta = 0$	Prob>  T
INTERCEP	23.807	0.0001
X2	-13.055	0.0001

independent variable to analyze. The income for the cities included in the research is the 1999 incomes posted at the website:

[verticals.yahoo.com/cities/categories/medfamily.html](http://verticals.yahoo.com/cities/categories/medfamily.html)

which is being used as the best available surrogate.

Using multiple linear regression to include price elasticity, cross price elasticity, and income elasticity we have the three fare product market demand functions listed above. The  $r^2$  and  $r^2$  (adj) as compared to our original three fare product market demand functions did not increase significantly. It is proposed that income elasticity should be dropped from the model and thus not included in the optimality procedure. We can quickly verify the significance, if any, that income may have on  $q_1$  by observing the sums of squares of regression between the three independent variables  $x_1$ ,  $x_2$ , and  $I$ .

Now, including the income elasticity we have the following results from the ANOVA tables for  $q_1$ :

$$SSR(I|x_1, x_2) = .70009, \quad r_{I|x_1, x_2}^2 = .070178$$

Additionally, we have the following variance inflation for the three independent variables and their p-values from the above listed ANOVA tables for  $q_1$ :

$$x_1 : 4.4024, \quad x_2 : 3.9922, \quad I : 1.0172, \quad p_{x_1} = .0001, \quad p_{x_2} = .0001, \quad p_I = .0966.$$

The p-value is the probability for the t-test. The p-value is too high for the independent variable income. The F-partial test also agrees with these results.

Table 3: Analysis of Variance:  $q_1$  regress on  $I$ 

Source	DF	Sum Sq.s	Mean Sq.	F-Value	Prob> F
Model	1	1.59173	1.59173	2.782	0.0966
Error	250	143.01307	0.57205		
C Total	251	144.60480			

Root MSE 0.75634      R-square 0.0110      Adj R-sq 0.0071

## Parameter Estimates

Variable	DF	$\beta$ Estimate	Standard Error
INTERCEP	1	-2.574539	4.04070019
I	1	.650506	.38997346

  

Variable	T: $\beta = 0$	Prob>  T
INTERCEP	-.637	.5246
I	1.668	.0966

Thus income does not make up for the unexplained variation. Observation of the ANOVA tables for  $q_2$  and  $q_3$  can be done in the same way and yields similar results. Thus income elasticity is not a necessary variable for any of the three market demand functions and shall be removed from the model.

The fact still remains that the  $r^2$  for  $q_1$ ,  $q_2$ , and  $q_3$  are .411, .411, and .468 respectively for the sample of 250 flights when  $I$  is excluded. Questions arise for the improvement of the model: (1) There might be some terms we should include in the model that can help explain the proportion of variations. For this consideration, an immediate improvement on the model would be to include cross product terms of  $x_i$  and  $x_j$  in the model of  $q$ . This implies considering the family of the exponential models for  $Q$  and a flexible functional form would be the translog model (see Final Remarks). (2) Do the unusual observations have such a large effect on the variation? If we exclude a few of the unusual observations, or place less weight on these observations, the variance is nearly constant for each of the three fare product demands. Further analysis of the plot of  $q_1$  versus the predicted value for  $q_1$ , if we exclude a few unusual observations then  $r^2 = .5023$ . Further analysis reveals these unusual observations were data from flights in the northeastern states whose prices for first class and standard economy class were extremely high. Several flights had first class prices above \$1200 and standard economy class above \$900. So further research is needed to improve the model. Research that may involve looking for a key indicator, possibly for the regions of the 48 states since there is some variation in prices

Table 4: Analysis of Variance:  $q_1$  regress on  $x_1, x_2$

Source	DF	Sum Sq.s	Mean Sq.	F-Value	Prob> F
Model	2	59.416	29.708	86.83	0.0001
Error	249	85.189	.342		
C Total	251	144.60480			

Root MSE 0.5849    R-square 0.4141    Adj R-sq 0.406

Parameter Estimates

Variable	DF	$\beta$ Estimate	Standard Error
INTERCEP	1	9.6117	.4804
X1	1	-.2533	.1658
X2	1	-.6503	.1467

Variable	T: $\beta = 0$	Prob>  T
INTERCEP	20.01	0.0001
X1	-1.53	0.128
X2	-4.43	0.0001

in the different regions.

From the analysis of the market demand functions we have no reason to doubt our normality assumptions and no reason to doubt our optimality model shall exclude income elasticities. These market demand functions reveal consumer behavior within these 250 flights. It is obvious that the changes in price of standard economy class fare products directly effects demand for first class and discount fare class. We can observe the effects of consumer behavior; that is passenger diversion, from these market demand functions. These cross price elasticities offer a clearer picture of the number of passengers who will divert to a lower priced fare product given an increase in price. From the multiple linear regression model, we have observed how consumers react to the degradation costs and the passenger diversion that occurs once we increase or decrease a price of other fare products. Forecasting the future behavior of passenger diversion based on their current behavior is desired.

## 4 Optimality

The objective now is to maximize revenue, where the decision variables are the prices of the three fare products. Recall,  $x_i$  is the natural logarithm of the price of fare product  $i$  and  $q_i$  is the natural logarithm of the quantity demanded for

fare product  $i$ . Initially our objective function defined in terms of price and quantity yields the following problem.

$$\max R = \sum_{i=1}^3 e^{x_i} e^{q_i},$$

subject to

$$\sum_{i=1}^3 e^{q_i} \leq \text{capacity}.$$

However, to analyze revenue in terms of price only, we shall rewrite the objective function to include the values of  $q_i$  in terms of  $x_i$ . Also the constraint shall be rewritten in terms of price. For linearity purposes, the model is designed such that the input for capacity ( $CAP$ ) will be the logarithm of the capacity of the aircraft. Therefore we have the following problem.

$$\max R = e^{9.61+.747x_1-.650x_2} + e^{10.8+.45x_2-.224x_1-.123x_3} + e^{10.7+.828x_3-.793x_2},$$

subject to:

$$31.11 - .477x_1 - 1.99x_2 - 0.295x_3 - CAP \leq 0.$$

Therefore, we have a nonlinear optimization problem of the constrained case.

Our objective function is clearly convex. However, the objective function is bounded by the capacity of the aircraft. The optimal prices for revenue shall occur when the  $\sum_{i=1}^3 e^{q_i}$  is exactly equal to the capacity of the aircraft. Therefore, to find the optimal revenue we apply the Lagrange multiplier method to solve for optimal prices using price elasticity and cross price elasticity as our independent variables.

Our objective function is

$$f(x_1, x_2, x_3) = e^{9.61+.747x_1-.650x_2} + e^{10.8+.45x_2-.224x_1-.123x_3} + e^{10.7+.828x_3-.793x_2}.$$

The constraint is:

$$g(x_1, x_2, x_3) = 31.11 - CAP - .477x_1 - 1.99x_2 - .295x_3 = 0.$$

From the condition  $\nabla f = \lambda \nabla g$ , We can solve the unknowns  $x_1, x_2, x_3$  and  $\lambda$ , and thus, the optimal prices for revenue.

The data used for the econometric modeling was based on the average daily purchases for the flights. This model is constructed such that an input value for the daily capacity for the aircraft would yield optimal prices for revenue given the market demand functions constructed had a larger  $r^2$ . Since the values of  $x_i$  is  $\ln P_i$ , there is a large difference between an  $x_1 = 6.4$  and  $x_1 = 6.46$ , a difference of 38 dollars. Thus when solving the system of equations, we must be very careful to watch the precisions of the digits.

This model appears to be a sound method to use. The model is set up to maximize daily revenues based on price and cross price elasticities. The steps for analysis are clear and the structure of the optimality is clear. Further research into indicator variables for the market demand functions could yield a more accurate model and then using the optimality equations in the same way should prove to output more realistic prices for each of the three fareclasses. The study could be extended to include international and domestic flights. Then the model would have up to N-fare products. The structure would be the same, the market demand functions would be as follows:

$$\hat{q}_i = \hat{\beta}_{i0} + \hat{\beta}_i x_i + \hat{\beta}_{i(i-1)} x_{i-1} + \hat{\beta}_{i(i+1)} x_{i+1}.$$

And we would like to

$$\max R = \sum_{i=1}^N e^{\hat{\beta}_{i0} + (1+\hat{\beta}_i)x_i + \hat{\beta}_{i(i-1)}x_{i-1} + \hat{\beta}_{i(i+1)}x_{i+1}},$$

subject to

$$\sum_{j=1}^N (\hat{\beta}_{j0} + \hat{\beta}_{j1}x_1 + \hat{\beta}_{j2}x_2 + \dots + \hat{\beta}_{jN}x_N) - CAP = 0.$$

## 5 Final Remarks

1. Despite its simplicity, the linear model is too restrictive and cannot accommodate for the variation in the data. Modern studies of demand and production are usually done in the context of a flexible functional form. Flexible functional forms are used in econometrics because they allow analysts to model second order effects. The most popular flexible functional form is the translog model, which is often interpreted as a second-order approximation to an unknown functional form. Let  $\ln y = f(\ln x_1, \dots, \ln x_k)$ . Then its second-order Taylor series around  $(x_1, \dots, x_k) = (1, \dots, 1)$  is in the form

$$\ln y = \beta_0 + \sum_{i=1}^k \beta_k \ln x_i + \frac{1}{2} \sum_{i,j=1}^k \gamma_{ij} \ln x_i \ln x_j + \epsilon.$$

Since the value of  $r^2$  in the linear model is at best 0.473, we may consider to apply the translog model for a better fitting. Recently, translog models were used in [3] for the study of productivity change model in the airline industry.

2. Principal component analysis (PCA) involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and

each succeeding component accounts for as much of the remaining variability as possible. In principal component analysis (PCA), the data are fit to a linear model by computing the best linear approximation in the sense of the quadratic error.

Have noted that the  $r^2$  could be improved to a more satisfactory level, we may want to include more variables in the linear model. However, to select as few key independent variables as possible in the model, applying the PCA technique in the modeling will be a good idea. Recently, PCA was applied in [1] for the evaluation of deregulated airline networks with an application to Western Europe.

3. In [5], a new analytical procedure for joint pricing and seat allocation problem was developed using polyhedral graph theoretical approach considering demand forecasts, number of fare classes, and aircraft capacities. Three equivalent models were formulated in the paper: The first model is a 01 integer programming model. The second model is obtained from the first model using the notion of constraint aggregation. The third model is derived by exploiting the special data structure of the first model and utilizing the concepts of split graphs and cutting planes. A decision-support tool was developed for price structure designers to be able to consider a wide variety of possibilities concerning the number of fare classes.

**Acknowledgments:** This work was supported in part by NSF-IGMS (0408086 and 0552377) and MTSU Research Enhancement Program for Hong.

## References

- [1] N. Adler and B. Golany, Evaluation of deregulated airline networks using data envelopment analysis combining with principal component analysis with an application to Western Europe, *European Journal of Operational Research*, **132** (2001), 260–273.
- [2] T.C. Botimer and P.P. Belobaba, Airline pricing and fare product differentiation: A new theoretical framework, *Journal of the Operational Research Society* **50** (1999), 1085–1097.
- [3] R. Ceha and H. Ohta, Productivity change model in the airline industry: A parametric approach, *European Journal of Operation Research* **121** (2000), 641–655.
- [4] N.R. Draper and H. Smith, *Applied Regression Analysis*, Third edition, John Wiley and Sons, Inc., New York, 1998.
- [5] A. Kuyumcu and A. Garcia-Diaz, A polyhedral graph theory approach to revenue management in the airline industry, *Computers and Industrial Engineering*, **38** (2000), 375–395.

- [6] S.R. Morrison and C. Winston, *U.S. Aviation After 20 Years of Deregulation*, Brookings Institution: Washington D.C., 1995.
- [7] D.L. Rubinfeld and R.S. Pindyck, *Econometric Models and Econometric Forecasts*, McGraw-Hill, Massachusetts, 1998.