Estimation of Demand and Supply in US Airline Industry

Ran Wang
Georgia Institute of Technology
Email: wangran2010@gmail.com

Weiwei Yang
SunTrust Bank
Email: wwy151430@gmail.com

Abstract
Airline ticket prices are simultaneously determined by the supply and demand. Airlines could make more profit if they know the factors that influence the demand and supply. For the consumers, they would really higher utility if they know the pricing strategies of airlines. In this study, we analyze the factors that drive the demand and supply in the airline industry. Using the post-sale ticket level data, we are able to identify the potential factors. In this way, we shed light on future pricing strategies for the airlines and future purchase behavior for the consumers.

Keywords: Demand; Supply; Pricing Strategies; Consumer Behavior

JEL codes: L11, L22

Introduction

Airline ticket pricing is often simultaneously determined by the supply and demand. Supply demand mismatch in this industry often leads of sub-optimal profits for the airline firms and unsatisfactory utilities for the consumers. Airline ticket pricing is in a constant state of flux. We often get to experience its dynamic nature first hand. Hence, studying about the factors that affect the supply and demand in the airline industry got us excited. In this paper, we intend to study the factors that affect both the supply and demand for the airline industry.

Our first step in this paper was data collection and manipulation. We identified the data that we would need for this study. We made extensive use of the material available from the Department of Transportation and the Bureau of Labor Economics to obtain the data. We further filtered and transformed the data in order to obtain the variables that were used in our models. Then, we built static models for Supply and Demand using Multiple Linear Regression and Two Stage Least Square Regression respectively. These models were helpful in understanding how the supply and demand behave with respect to each other. Finally, we used a dynamic model including system equations to capture the interaction between the supply and the demand, which would be helpful in optimizing supply and demand at the same time.

The reason that we include both static model and dynamic model is due to the asymmetry of information. Previously, the firm did not know much about consumers' utility and consumers did not know much about pricing strategy for airline firms. They made their decisions based solely on their knowledge. In such a situation, the Static model is suitable to capture the pricing behavior of airline firms and purchasing behavior for the consumers. However, nowadays, more and more information from consumers are exposed to airlines and airlines' pricing strategy is analyzed in many studies. The information is more transparent and symmetric presently. More and more interactions between demand and supply are available now. Under this situation, it would be more reasonable to apply dynamic model to fit both demand and supply simultaneously. So, this paper presents the estimation results based on both static and dynamic model and further comparisons are made between them.

There are a lot of empirical studies that discuss the pricing strategies and pricing issues. Borenstein and Rose (1994) and Gerardi and Shapiro (2009) talks about the relationship between airline competition and price dispersion. Brueckner, Dyer and Spiller (1992) show the factors that affect airline pricing with hub-and-spoke networks. Chakrabarty and Kutlu (2014) talks about how the competition affect the price discrimination in the airline market. Hazledine (2006) illustrates price discrimination under Cournot market. Wang, Han, Zhang and Zhang (2009) analyze the factors that influence pricing behavior for the big supermarkets in Tianjin.

This study could be very useful to several entities. The findings of this study can be of use to the airline firms. By understanding the factors that affect the supply and demand for the tickets will enable them to make better decisions regarding the pricing and quantity for different routes. This study can also help the consumers in maximizing their utility. The lesser the price of the ticket they purchase, the higher their utility. We believe that the government, to make decisions regarding its policies could use this study. Often government intervenes and makes policies in special situations. An example of such situation is when the oil prices become extremely high. They refer to the demand supply parameters to make their decisions.

The rest of the paper is organized as follows. Section II explained the data that is used in this analysis. Section III described the demand and supply model and model specification. Section IV discussed the model results. The last section summarized the main findings in this study.

The data

Data Collection and Manipulation constituted a significant portion of our study. We began with understanding the problem, and researching about the factors that could have an effect of the supply and demand in the airline industry. There are three categories of data: quantity and price data; cost price data for the airlines and demographic data. We decided to use four quarters of year 2009 for this project, as we believed it will be a good indicator of the present times. Also the data for all the variables is available for year 2009.

Quantity and price data are obtained from Passenger Origin-Destination Survey of the U.S. Department of Transportation (DB1B data set). This data set is a 10% random sample of all tickets that originate in the United States on domestic flights. Now, we list some of details for the data construction process. First, all multi-destination tickets are dropped, as it is difficult to identify the ticket's origin and destination without knowing the exact purpose of the trip. Second, any itinerary that involves international flights is eliminated. Third, we adjust the fare class for high-end carrier. That is, for some airlines, due to marketing strategies, only first class and business class tickets are provided to consumers on all routes, especially some small airlines. However, the quality should be taken as coach class. So, in our project, we consider all such tickets as coach class tickets. In different time periods, due to the pricing strategy changes, sometimes high-end-only carrier switches to a regular carrier, which sells both coach class tickets and high-end tickets. So, we treat the tickets in each quarter separately when considering the adjustment. Fourth, tickets that have high-end segments and unknown fare class are dropped. We then proceeded to generate the quantity variable. It denoted the total number of tickets that we purchased for each airline, for each route in each quarter. We also generated a variable called stage length, which was the average distance that was covered by the planes for each route in each quarter. This was an indication of the quality of the airline. Moreover, distance variable is generated as the shortest directional flight distance. Lastly, we introduced the online rate variable, which was the inverse of the direct flight rate. We averaged this quantity for each airline, each route and each quarter.

All tickets with incredible prices are dropped from our data set. We eliminated the open-jaw tickets since it would be difficult to distribute the ticket price into outbound and inbound segment for open jaw tickets. We eliminated the tickets that have a price less than 25 dollars or higher than 99% percentile or more than 2.5 times deviation from mean for each airline within a route. The tickets that have price less than 25 dollars are considered as frequent flyer program tickets and the tickets that have prices higher than 99 percentile are considered to be input (key punch) errors for the data set. For the round trip tickets, we divided the total price by two to get the one-way price. Finally, from the DB1B data set, we constructed the price for each itinerary, the average price for each route in each time period for coach class.

The cost data set is constructed from the firm level data of DOT's airline production data set (based on Form 41 and T100). Labor cost and energy cost are obtained from these two data sets. The salaries and benefits for five types of personnel are provided in Form 41/P6. Annual employee numbers are given in Form 41, P10. We interpolated the annual employee data to get the quarterly values. Based on total salary and benefit data and total employee data, we generate a labor price index as the average per employee salary and benefit. From Form 41/P12, we get the total domestic fuel cost and total fuel consumed (gallons) for each quarter. We derive the per gallon fuel cost as a fuel price index. Note that all these cost shifters are firm level, not route-firm level.

In order to estimate the demand, we also include the city specific demographic variables: per capita income (PCI) and population (POP). We get the city level PCI and population data from Bureau of Labor Statistics. We interpolate the annual data to get the quarterly PCI and population for each city. For each origin-destination city-pair, we use the population weighted PCI as route-specific PCI index. Since only the metropolitan areas have the demographic information but some airports are located in small cities, this greatly reduces the number of the city-pairs in our final database.

Finally, we included two dummy variables to indicate if the airline carrier was a low cost carrier (LCC), Legacy carrier (big airlines), or neither. After exploratory analysis of the variables that we generated, we found that we can get better linearity by log transformations. Hence, we transformed all the quantitative variables accordingly.

Static Model

Our initial model was based on the assumption that the demand and supply were static. Under the static model, we assume that the information between airlines and consumers are not symmetric. That is, the consumers only know the utility maximization structure and their preference. So, the consumers only make decision on purchase based on their preference. In this circumstance, the consumers do not know any pricing strategies from airline firms. On the other side, airline firms do not know the consumers utility structure, so they only make pricing strategies based on profit maximization with consumer side given. We later proceeded to build a more complex system equation that took into consideration interaction between supply and demand. Under this condition, the firms know consumers' preference. Also consumers know the firms' pricing strategies. The interaction between consumers and airlines makes the pricing strategy and purchasing strategy is based on full information. That is, consumers know how the airlines will react to their demand elasticity, so they do not treat the pricing strategy as given. Also, the firms know that the consumers will be more or less sensitive to the firms' price due to their knowledge of firms' pricing strategies. In this way, the airline firms treat the consumers' purchase decision as dynamic, responsive to firms' pricing. Under this circumstance, the consumers and firms make their decisions simultaneously. The dynamic system model is introduced to get the demand and supply elasticity. In the following part, we will show how we develop the static models and dynamic model.

Demand Model

In this model, the response variable is total quantity that the consumers consumed, demand for airline tickets. We control for many factors that might influence the demand quantity. The predictors here are price for each firm on each route in each time period, price (p), the population (pop), per capita income (pci), distance (dist), stage length(sl), and Low Cost Carrier and Legacy airline dummies. Also we control quarter dummies.

The reason that we control for price is that from demand theory, higher price means lower demand. Consumers' direct utility level is greatly influenced by the price of the tickets. Demographic factors, such as population and per capita income are the most important two factors that would influence the demand for each route greatly. We expect that larger cities with more population will definitely have larger demand. Also per capita income influences the highest willingness to pay, which shift up the demand curve and demand quantity. Online rate is defined as the inverse of change of plane rate. To be specific, for one specific ticket, online rate is equal to 1 when there is no change of plane and 0 when there is change of plane. The reason that changes of plane matters is that more change of planes leads to higher possibility of loss of package and thus more inconvenience. As other variables, we aggregate all ticket level data into firm level data to get average characteristics. Online rate should negatively influence the demand side. Stage length is the average stage length that the airline flies for each segment. Distance is the minimum directional flight distance for given route. Our model includes both distance and stage length in demand model such as the stage length means service quality. Longer stage length means longer flight time, so the consumers would spend more time on airplane. It brings some time waste for the consumers. However, higher stage length also leads to accumulation of more frequent flyer miles, which benefits the consumers. So, it brings both positive and negative effects on demand. Low Cost carrier and Legacy carrier dummies are included here to control for the difference between large airlines and small local airlines. Seasonality also matters because it correlates with temperature, which greatly influences the tourism routes' demand. In summary, we include all factors that might influence the demand in full model and carry on model selection method to select the model that explain our model best.

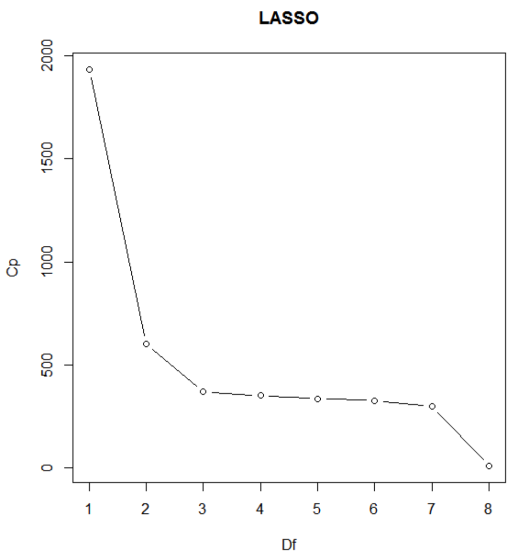
We proceeded with the exploratory data analysis of the variables. This convinced us that a multiple linear model could be a good fit for our scenario.

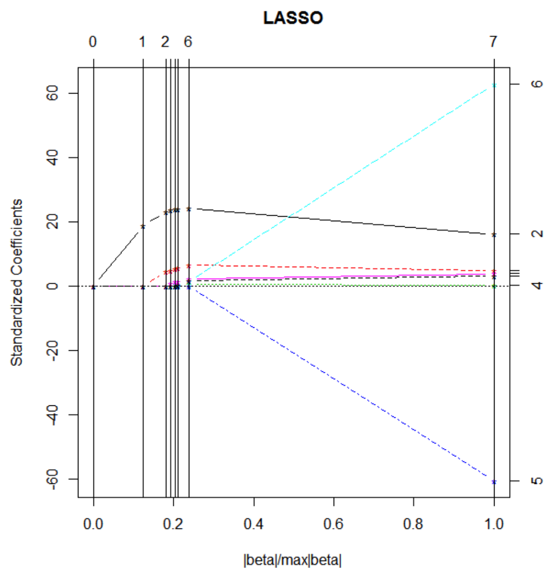
The full model for static demand model is given as

$$q = \beta 0+\beta 1p + \beta 2pop + \beta 3pci + \beta 4online + \beta 5sl + \beta 6dist + \beta 7lcc + \beta 8legacy + \beta 9qtr1 + \beta 10qtr2 + \beta 11qtr3 + \varepsilon$$

The next step in our analysis was model selection. We used the following three methods for model selection: exhaustive search, step-wise regression (forward, backward, both) and Lasso Regression.

The following plot shows our results from Lasso Regression:





The results that we obtained from all the models were similar. We saw that all the variables that we included were statistically significant, hence we chose to include all of them in our model (as shown in the lasso plot). The estimation results from this model can be found in the Table 1.

Supply Model (Static)

For the supply side, the firms will decide the price for the airline tickets based on the supply cost and some miscellaneous factors mentioned below. The difference here is that the firms take the supply cost into consideration. Labor cost is captured by "labor", which is the average per employee salary and fuel cost is captured by "fuel", which is the per gallon fuel cost. All other variables are following the same definition as demand model. Higher supply cost leads to higher supply price. That is why we control for these factors.

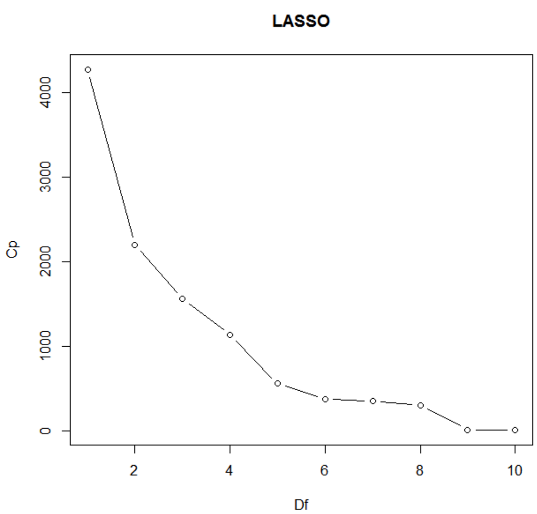
The most important difference here is that the quantity variable is endogenous variable in supply model because many other factors (included in error term) also influence the supply price and these factors also correlated with quantity, causing the total supply estimation to be biased. However, due to the limited data, we cannot capture all the factors that would both influence the price and quantity. Here we introduce one instrument, quantity from all other routes that share the same origin to solve the endogeneity problem. The reason that the quantity from all other routes serves as potential good instrument is that it controlled for all other factors that might influence the quantity. However, after controlling for this variable, the quantity does not correlate with the error term any more. Two stage least square regression is used here to include the instrument into supply model.

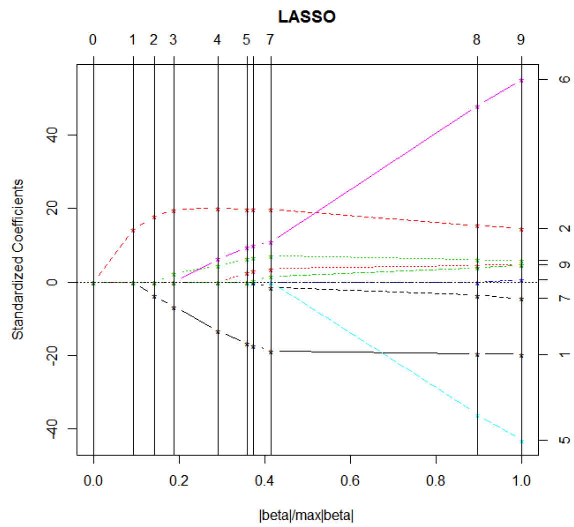
We proceeded in a manner similar to the Demand Model. We performed the exploratory data analysis. The full model for static supply model is given as

$$p = \beta 0 + \beta 1 q + \beta 2 dist + \beta 3 sl + \beta 4 labor + \beta 5 fuel + \beta 6 lcc + \beta 7 legacy + \varepsilon s$$

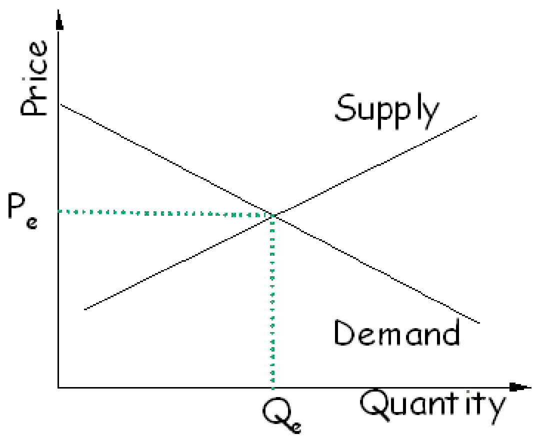
We then carried performed model selection using the following techniques: Exhaustive search, Step-wise regression (forward, backward, both) and Lasso Regression

All variables were statistically significant according to our model selection. The Lasso plot shown below confirms it. The estimation results for this model are shown in Table 2.





Demand and Supply system Model (Dynamic Model)



Representative figure of Demand Supply Dynamic equation

In the dynamic model, the consumers make purchasing decisions based on full information about firms' pricing strategy and firms make pricing strategies based on full information about consumers' utility structure. Consumers and firms make decision simultaneously, so the system equation is introduced here. The system equation just combines the demand mode and supply model together to build up a demand and supply equation system, which allows for interactions between firms and consumers. Also, we use the same instrument as in supply model. The equation system is given as

$$q = \beta 0+\beta 1p + \beta 2pop + \beta 3pci + \beta 4online + \beta 5sl + \beta 6dist + \beta 7lcc + \beta 8legacy + \beta 9qtr1 + \beta 10qtr2 + \beta 11qtr3 + \varepsilon$$
$$p = \beta 0+\beta 1q + \beta 2dist + \beta 3sl + \beta 4labor + \beta 5fuel + \beta 6lcc+\beta 7legacy+\varepsilon s$$

To carry on the estimation, three stage least square is introduced here. The estimation results for static demand model and dynamic model for demand part is given in Table 1.

Results and Interpretation

Table 1 shows the estimation results and numerical comparisons between static and dynamic demand model.

Table 1. Comparison between Static Demand Model and Dynamic model

Quantity	Static Demand	Dynamic Model
Price	-1.643***	-2.954***
	(-71.58)	(-26.64)
Stage Length	2.841***	2.952***
	-70.31	-69.5
Distance	-2.691***	-2.487***
	(-63.74)	(-49.79)
Online	0.250***	0.218***
	-6.44	-5.09
Population	0.583***	0.579***
	-46.53	-45.07
Per Capita Income	2.694***	2.755***
	-40.35	-37.64
Low Cost Carrier	-0.0905**	0.0792*
	(-3.23)	-2.43
Legacy Carrier	-0.0725*	0.492***
	(-2.48)	-9.29
Quarter 1	-0.243***	-0.262***
	(-13.84)	(-14.52)
Quarter 2	-0.137***	-0.171***
	(-7.95)	(-9.08)
Quarter 3	-0.015	-0.0342
	(-0.87)	(-1.94)
Constant	-20.52***	-17.38***
	(-32.18)	(-26.60)
N	17865	17865

Note:

1. t statistics in parentheses and * $p<0.05$, ** $p<0.01$, *** $p<0.001$.
2. Quantity is endogenous variable, so we use quantity from other routes that originate from the same route as an instrument.

The following interpretations and explanations can be made for the demand based on our study. If the ticket price increases, less people choose to take plane; they will switch to using automobile for short trip. Consumers prefer longer stage length since it brings more mileage credit for frequent fly membership even though it also brings about the longer time wastage. Consumers prefer no change of plane since online rate have positive effect in both models. Routes with more population and higher income have more demand. Low cost carrier and legacy carrier have effect on demand. Season is important especially for tourism purpose trip. When we compare the two results, we can see that in the dynamic interaction framework, demand is more elastic. So, in the full information setting, the consumers are more sensitive to firms' price when they know firms' pricing strategy. Also other variables' effect is different. Since they are not so important, we will not explain their differences here. From the differences, we show how the information changes the consumers' behavior. That is exactly why the consumers' survey works to improve the social welfare.

The estimation results for static supply model and dynamic model for supply part is given in Table 2.

Table 2. Comparison between Static Supply Model and Dynamic model

Price	Static Supply Model	Dynamic Model
Quantity	-0.0505***	-0.0139***
	(-9.15)	(-3.52)
Labor cost	0.235***	0.204***
	-24.08	-19.78
Fuel cost	-0.0123	0.000189
	(-1.04)	-0.02
Stage length	0.304***	0.147***
	-11.77	-7.26
Distance	-0.0758**	0.0894***
	(-2.79)	-4.19
Low Cost Carrier	0.0378***	0.0574***
	-4.1	-5.87
Legacy Carrier	0.356***	0.387***
	-39	-41.47
Constant	2.777***	2.452***
	-39.55	-44.33
N	17856	17856

Note:

1. t statistics in parentheses and * p<0.05, ** p<0.01, *** p<0.001.
2. Quantity is endogenous variable, so we use quantity from other routes that originate from the same route as an instrument.

The following interpretations and explanations can be made for the Supply based on our study. Quantity measures tradeoff between demand effect and scale of economies effect. That is, for large demand the price should be higher. However, larger demand means lower average cost, and thus lower supply price. For our case, the scale of economies effect dominates. So, higher quantity, in total, have lower supply price. Labor cost and fuel cost shifts up the supply price. The effect from labor cost is obvious in both models. However, fuel cost is not so significant in both cases. On one side, fuel cost shifts up the supply cost. On the other hand, higher fuel cost also indicates that for more consumers will switch to airline due to higher fuel cost associated with automobile trips. So, the fuel cost, kind of, captures the two conflicting effect, leading to insignificant results. Difference between static supply model and dynamic model shows how the information changes the firms' behavior.

Conclusion

This paper built up a static and dynamic model for the airline industry. Using the post-sale data from 2009, this paper analyzed the factors that influence demand and supply in the airline industry. In the static models, we assume that the information between the consumer and the airlines is not shared, where as in the dynamic models, we assumed that all the information is shared between the consumer and the airlines. This study can have direct applications in the airline industry, government regulations and consumer satisfaction.

This papers found consumers prefer longer stage length, no change of plane. Additionally, routes with more population and higher income have more demand. Low cost carrier and legacy carrier have effect on demand. Seasonality is found in this analysis. The scale of economies was found for the airline industry, that is, airlines with higher supply have lower supply price. The differences between static model and dynamic model suggest the potential changes in the consumer behavior and production decisions if the information becomes more symmetric.

References

Borenstein, S. and Rose, N.L. (1994), Competition and Price Dispersion in the U.S. Airline Industry, *Journal of Political Economy*, 102, 653-683.

Brueckner, J., Dyer, N., Spiller, P.T. (1992), Fare Determination in Airline Hub-and-Spoke Networks, *RAND Journal of Economics*, 23, 309-333.

Chakrabarty, D. and Kutlu, L. (2014), Competition and Price Discrimination in the Airline Market, *Applied Economics*, 46, 3421-3436.

Gerardi, K.S., and Shapiro, A.H. (2009), Does Competition Reduce Price Dis-persion? New Evidence from the Airline Industry, *Journal of Political Economy*, 107, 1-37.

Hazledine, T. (2006), Price Discrimination in Cournot-Nash Oligopoly, *Economics Letters*, 93, 413-20.

Kutlu, L. and Wang. R.(2016), Price Dispersion, Competition, and Efficiency in the U.S. Airline Industry. Working Paper.

Kutlu, L. and Wang. R.(2016), Estimation of Cost Efficiency without Cost Data. Working Paper.

Wang. R. (2016), Revenue Efficiency and Dynamic Pricing in AirlineIndustry. Working Paper.

Wang, R., Han, N, Zhang, Y., and Zhang P., (2009),What influences pricing behavior of foreign supermarkets in Tianjin? *Management Observer*, 34, 58-61.

Wang, R. and Yang, W. (2016), Investment in anti-counterfeit label for branded firm, *Eurasian Journal of Economics and Finance*, 4(3), pp 43-53.

Wang, R. and Yang, W. (2016), What influences the pricing of automotive in U.S. market?*International Journal of Liberal Arts and Social Science*, 4(5), pp. 88-97.

Wang, R. and Yang, W. (2016),How CAFÉ influences auto firms and consumers? Using evidence from U.S. market, *Eurasian Journal of Economics and Finance*, 4(3), pp. 85-99.