Title of submission:

"Exploring Airline Fare Pricing"

Author:

Keith A. Willoughby
Department of Finance and Management Science
Edwards School of Business
University of Saskatchewan
Saskatoon, SK Canada S7N 5A7
Phone: (306) 966-2128
Fax: (306) 966-2515
E-mail: willoughby@edwards.usask.ca

KEY WORDS: Regression analysis; Airlines

Abstract:

We present a quantitative model to investigate the pricing of airline fares. As a particular case study, we examine fares offered in 2003 and 2004 at several competing airports in the U.S. northeast. Our selection of airports in this region permits us to examine two important issues. First, we explore the price competitiveness (or lack of same) of some smaller, regional airports. Second, we illustrate the effect of Southwest Airlines' recent entry into Philadelphia on airline fares offered from that city. By comparing our statistical results prior to and after Southwest's entry, we note the growing attractiveness of Philadelphia airfares.

1. Introduction and Literature Review

The U.S. Department of Transportation's Bureau of Transportation Statistics (2001) compiles an annual record of airport activity statistics. Among the many pieces of data it collects, it reports the annual total number of enplaned revenue passengers. In 2000, nearly 640 million passengers traveled by airplane. In 2001 (perhaps due to heightened air travel fears due to the September 11th tragedies), this number fell to roughly 595 million, a drop of 6.8%. Indeed, this marked the first time since 1991 that the number of enplaned revenue passengers had declined. From 1992 to 2000, the average annual growth rate in total passengers was a healthy 4.6%.

We present a linear regression model to explore airline fare pricing offered at several airports of varying size in the U.S. northeast. We investigate five airports in Pennsylvania (Philadelphia, Pittsburgh, Harrisburg, State College and Williamsport) as well as one airport in Maryland (Baltimore). Such a selection of airports for this case study permits us to investigate the price attractiveness – or otherwise – offered at some smaller, regional airports. To wit, what is the extent of price premiums a traveler would pay to fly from a smaller airport? Further, what are the key independent variables (specific city of origin, flight mileage, population of destination city, number of layovers) in explaining airline fare pricing?

Besides these objectives, our model results will allow us to determine how the recent entry of a discount airline operating from Philadelphia (in this case, Southwest Airlines) has affected fares offered at that particular airport. By conducting a multi-year analysis (as opposed to a single "snapshot" of airline fares at one point in time), we may be able to observe changes in fare competitiveness. We can explore this effect by conducting separate regression analyses on fares before (2003) and after (2004) Southwest's entry.

The emergence of discount, no-frills airlines has certainly given the price-conscious traveler alternative options to consider when desiring to fly. Southwest Airlines, in particular, has carved a profitable niche in the U.S. airline industry. Founded in 1971 by Herb Kelleher and Rollin King, it has become the nation's largest carrier in terms of customers boarded. It began with service between Dallas, Houston and San Antonio, but now operates in 59 U.S. cities in 31 states, with Phoenix, Las Vegas and Baltimore being its three busiest airports. In 1995, it became the first airline to launch its own website and in 1998, Fortune Magazine named it as the best place to work in America. Currently, it carries over 65 million passengers annually on about 2,800 flights each day with a fleet of roughly 400 Boeing 737's (www.southwest.com).

Southwest's original strategy of short-haul but high frequency flights, combined with its ongoing commitment to lower fares, has contributed to its remarkable success. Its foray into a particular market does not go unnoticed by its competitors, the traditional "legacy" airlines. For example, Lin, Dresner and Windle (2001) document that the entry of Southwest Airlines into a market causes the full-service airlines to dramatically reduce their fares in order to maintain some semblance of competitiveness. Southwest appears to have an effect on its competitors.

As the reader may be unfamiliar with the size and operations of the three smaller airports represented in our case study, we provide some brief commentary. Harrisburg International Airport is the largest of the three. It operates more than 120 flights each day, with 7 airlines flying to 14 destinations (13 domestic and 1 international). The State College airport includes 4 airlines operating 21 flights daily, while Williamsport is served by a single airline (U.S. Airways) with 5 flights on weekdays, 2 on Saturday and 4 on Sunday.

Since Southwest Airlines does not operate flights from these three airports, those travelers wishing to take advantage of Southwest's competitive fares must look elsewhere for bargains. (In fact, up until May, 2004, Southwest failed to serve any communities within the state of Pennsylvania, suggesting that price-conscious customers may be inclined to travel out of state for cheaper fares). Certainly, price may not be the only factor influencing a traveler's airport selection decision (for example, Windle and Dresner (1995) use a logistic model to show that airport access time and flight frequencies are important predictors of airport choice). However, one cannot completely ignore the effect of price on one's decision, especially in this age of price-savvy customers using web-based resources to continually search for the lowest fare.

In the next section of our paper, we describe the linear regression model. We provide results in the third section, while concluding remarks are offered in the final section.

2. Linear Regression Model

Our desire is to quantify the relationship between a response variable (airfare) and a number of explanatory variables (e.g. particular point of departure, mileage, population of origin, population of destination, and number of layovers). While it seems obvious that average airfares in much smaller communities ought to be higher, the magnitude of the relationship is not known. Linear regression can be used to provide an enhanced understanding of the quantitative relationships between groups of variables. Indeed, we can obtain a more in-depth insight than what can be gained by solely examining average values of a response variable.

Our linear regression model is of the form:

$$Y = \beta_0 + \beta_1 \text{ IPT} + \beta_2 \text{ SCE} + \beta_3 \text{ MDT} + \beta_4 \text{ PHL} + \beta_5 \text{ PIT} + \beta_6 \text{ Mileage}$$
$$+ \beta_7 \text{ PopLarge} + \beta_8 \text{ PopSmall} + \beta_9 \text{ OneStop} + \beta_{10} \text{ MultStops}$$

The labels for the first five explanatory variables (IPT through PIT) refer to the three-letter airport city code identifier used for that specific airport. These explanatory variables are actually dummy variables representing:

IPT = 1 if the particular flight departed from Williamsport
0 otherwise

SCE = 1 if the particular flight departed from State College
0 otherwise

MDT= 1 if the particular flight departed from Harrisburg
0 otherwise

PHL = 1 if the particular flight departed from Philadelphia
0 otherwise

PIT = 1 if the particular flight departed from Pittsburgh
0 otherwise

Mileage represents the air mileage between a specific origin-destination pair. Numerous internet sites list particular air mileage values (see, for example, http://www.webflyer.com/travel/milemarker/).

The explanatory variables PopLarge and PopSmall refer to the populations (in thousands) of the larger or smaller communities, respectively, in a specific origin-destination flight. We obtained the population values from the 2000 U.S. Census. We examined trips from each of our six origins (Baltimore, Williamsport, State College, Harrisburg, Philadelphia and Pittsburgh) to the 50 most populous cities in the U.S. (see the Appendix for a listing of these 50 cities). The populations of the three smaller communities, not obviously included in the nation's top 50, are as follows: Harrisburg (48,950), State College (38,420) and Williamsport (30,706).

OneStop and MultStops are explanatory variables used to represent the number of layovers in a specific flight. Being dummy variables, they have the following representation:

OneStop = 1 if a particular flight had exactly one stop
0 otherwise

MultStops = 1 if a particular flight had two or more stops
0 stop

Obviously, non-stop flights would be represented with values of 0 for each of these explanatory variables. We note that the constant term in our model (β_0) represents the airfare of a non-stop flight departing from Baltimore, controlling for the population of the cities in the origin-destination pair and total mileage.

3. Model Results

We obtained round-trip airfares in 2003 and 2004 from the Orbitz website, www.orbitz.com, one of the world's most popular travel websites (comScore Media Metrix recorded 11.1 million users who visited the Orbitz site at least once during the month of June, 2002). Although we recognize that airfares constantly change, we nonetheless had to arbitrarily select dates on which to retrieve data from this website. For the 2003 data, we gathered all of our airfares on July 7th and 8th, requesting Orbitz to provide us with the lowest fare (for an E-ticket, including all taxes and fees) between any two specific cities. Our hypothetical trips were scheduled to depart on Fri. October 17th and return on Tues. October 21st. In order to construct samples between the two years that mirrored each other as much as possible, we gathered our 2004 data for roughly the same time frame as in the preceding year. Our 2004 airfare data was gathered on July 5th through 7th for a hypothetical trip departing Fri. October 15th and returning Tues. October 19th. Admittedly, it is possible that our results may vary depending on the particular dates chosen for air travel, but we are confident in our approach for at least two reasons. Firstly, we gathered fare data for numerous trips, thus permitting us to observe any price differentials for specific points of departure across a wide spectrum of potential destinations. Secondly, we intentionally selected a departure date in the Fall period, ensuring that our trips would not be affected by late-summer seat-sales.

Incorporating 6 origins and 50 destinations provides a total of 300 origin-destination pairs. However, since Baltimore, Philadelphia and Pittsburgh are included in both our origin and destination lists (they are among the 50 most populous U.S. cities), we excluded these from our flight totals – in essence, we do not consider a flight from a city to itself! We also removed the Baltimore-Washington, D.C. flight pairing, since these two cities are very close to one another, and the Baltimore airport (better known as Baltimore-Washington International (BWI)) serves travelers from both communities. This left us with 296 origin-destination pairings.

We used Microsoft Excel to perform the regression analysis on data from two specific years. The 2003 data illustrates airline fares prior to Southwest's entry into Philadelphia, while the 2004 data portrays the fare situation subsequent to the discount airline's arrival in that city. Table 1 provides the results of our linear regression modeling for 2003, while Table 2 illustrates models results for 2004. The original model results are given in the 2nd column from the left in either table. For each variable, we list its unstandardized coefficient along with its standard error in parentheses. Many of the explanatory variables are highly significant. On the other hand, both of the population variables, PopLarge and PopSmall, appear to offer limited power in explaining airfares in our two-year sample.

Table 1
Linear Regression Model Results (2003)

Regression	Original	Revised	Long-Haul	Short-Haul
Constant	127.39 *** (18.32)	117.09 *** (12.13)	61.76 *** (21.21)	171.98 *** (16.90)
IPT	118.88 *** (16.60)	127.02 *** (12.05)	147.13 *** (15.05)	87.79 *** (17.72)
SCE	48.49 *** (16.59)	56.40 *** (12.24)	63.40 *** (15.20)	45.98 ** (18.68)
MDT	52.10 *** (16.38)	60.07 *** (11.97)	78.16 *** (15.20)	23.35 (16.84)
PHL	55.19 *** (12.63)	52.79 *** (12.16)	71.49 *** (15.76)	34.43 * (17.83)
PIT	36.78 *** (13.06)	40.29 *** (12.14)	43.39 *** (15.91)	42.37 ** (16.51)
Mileage	0.0805 *** (0.004)	0.0805 *** (0.004)	0.0873 *** (0.006)	0.0172 (0.025)
PopLarge	-0.00032 (0.003)	---	---	---
PopSmall	-0.01695 (0.0239)	---	---	---
OneStop	16.998 (10.668)	18.19 * (10.52)	45.74 *** (17.16)	12.14 (12.36)
MultStops	54.013 *** (11.702)	55.66 *** (11.42)	90.53 *** (17.57)	29.06 * (16.32)
R^2	0.720	0.719	0.701	0.299
Adjusted R^2	0.710	0.712	0.688	0.239
F	73.358	92.087	54.003	5.010
Significance levels: * p< 0.10 ** p< 0.05 *** p <.01				

Table 2
Linear Regression Model Results (2004)

Regression	Original	Revised	Long-Haul	Short-Haul
Constant	175.89 *** (15.88)	150.97 *** (10.99)	163.36 *** (21.66)	153.93 *** (14.39)
IPT	46.46 *** (15.76)	66.98 *** (11.87)	83.83 *** (15.65)	35.96 ** (15.50)
SCE	74.51 *** (15.79)	94.64 *** (12.06)	104.99 *** (15.90)	77.01 *** (15.77)
MDT	18.69 (15.21)	38.76 *** (11.37)	47.44 *** (15.09)	25.11 * (14.66)
PHL	-0.94 (11.61)	-8.09 (11.20)	-23.45 (14.55)	23.04 (14.94)
PIT	7.39 (11.87)	15.56 (11.23)	8.83 (14.95)	25.35 * (14.35)
Mileage	0.0656 *** (0.0039)	0.0659 *** (0.0039)	0.0654 *** (0.0056)	0.0437 ** (0.0211)
PopLarge	-0.0017 (0.0029)	---	---	---
PopSmall	-0.0427 * (0.0219)	---	---	---
OneStop	8.29 (9.72)	8.79 (9.76)	-4.59 (19.77)	24.63 ** (9.44)
MultStops	62.14 *** (12.42)	64.07 *** (12.45)	49.80 ** (21.68)	56.88 *** (16.48)
R^2	0.721	0.716	0.678	0.418
Adjusted R^2	0.711	0.708	0.664	0.368
F	73.554	90.428	48.441	8.425
Significance levels: * p< 0.10 ** p< 0.05 *** p <.01				

Due to the apparent insignificance of the population variables, we decided to drop them from any subsequent regression analyses. It may be that any population effects are somewhat captured in the respective originating airports (BWI, IPT, SCE, MDT, PHL or PIT) for each flight. We ran a revised linear regression model of the form:

$$Y = \beta_0 + \beta_1 \text{ IPT} + \beta_2 \text{ SCE} + \beta_3 \text{ MDT} + \beta_4 \text{ PHL} + \beta_5 \text{ PIT} + \beta_6 \text{ Mileage}$$
$$+ \beta_7 \text{ OneStop} + \beta_8 \text{ MultStops}$$

This data set may be accessed under the “Revised” sheet in either year’s spreadsheet. The results from this case are presented under the column heading “Revised” in Table 1 and Table 2. This regression model (more parsimonious than the original case) does a reasonably good job of explaining the variability in airline fares. The coefficients of determination are very similar in either year, with many of the explanatory variables being highly significant (p-values less than 0.01).

Interpreting the various coefficients allows us to more fully understand the pricing of airline fares in this region of the country. The constants in our models represent the fare for a non-stop flight from Baltimore, controlling for mileage. We note that almost all of the remaining airport dummy variables are positive, suggesting that price premiums (over and above Baltimore’s fare) are charged for flights originating from these locations. The one intriguing exception is Philadelphia in 2004. We note that in the first year of our data, its coefficient was positive; indeed, its value (52.79) was quite close to the coefficient values for smaller airports like Harrisburg and State College. In 2004, this coefficient became negative, signifying that Philadelphia may have become a better bargain for airline fares. Although its value was not overly significant, the fact that it was negative would appear to suggest that Southwest’s entry into the Philadelphia market has contributed to an overall decline in airline fares from this airport.

It seems that the competitiveness of the smaller airports in our case study experienced some volatility within our two-year sample. For example, in 2003, State College had a coefficient value (56.40) in the “ball-park” of other airports, even a large one like Philadelphia. In 2004, its coefficient value (94.64) was the highest of any of the airports, perhaps implying that its fares have become less attractive. Harrisburg, on the other hand, underwent an opposite trend. It had the second highest coefficient value in 2003. In the next year, its fares became somewhat more competitive as its coefficient value (38.76) was the lowest of our three smaller airports. We note that in 2004, its coefficient was still much larger than the corresponding values for Baltimore, Philadelphia and Pittsburgh.

To further document the pricing of airline fares, we considered each of our 50 destinations and determined the specific origin (of the six in our model) that provided the lowest airline fare for that destination. The results of our findings are illustrated in Table 3.

Table 3
Number of Instances of Least-cost Airline Fare

Origin	2003	2004	Change
Baltimore	23	19	-4
Williamsport	0	2	+2
State College	6	0	-6
Harrisburg	5	2	-3
Philadelphia	4	18	+14
Pittsburgh	12	9	-3
Total	50	50	

Unquestionably, Baltimore provides attractive bargains. It accounted for almost half of the least expensive fares in 2003, and about 40% of the best bargains in the subsequent year. As our linear regression results appeared to suggest, Philadelphia experienced a dramatic turn-around during our two-year sample. It went from providing the second least number of bargains in 2003 (only beating out Williamsport) to having almost the best number of least-cost airline fares in 2004. This may imply a further confirmation of the "Southwest effect"; namely, that the entry of this discount airline into a market encourages other airlines to reduce their fares. State College appeared to be a reasonably good choice for the price-conscious traveler in 2003 (it had the third most number of least-cost airline fares), but suffered a lack of competitiveness in the next year.

Frequently, flights are divided into long-haul versus short-haul trips. According to the U.S. Department of Transportation (1999), it defines a long-haul (short-haul) flight as any trip longer (shorter) than 750 miles. For purposes of this study, we shall observe the Department's threshold. We would like to understand the relationship between airlines fares, our specific variables, and whether the particular flight was long-haul or short-haul. By using the threshold value of 750 miles, we ended up with 193 long-haul flights and 103 short-haul trips. We then ran the revised multiple regression model for each of these two cases (see the long-haul and short-haul sheets in each of our data sets). The long-haul and short-haul results are given in their respective columns of Table 1 and Table 2.

The results from the long-haul flights mirror those from the revised model. Many of the variables are highly significant. In addition, the explanatory power of the long-haul cases matches the ability of the revised model to explain the variability in airline fares since the respective coefficients of determination are comparable. The only exception appears to involve the OneStop explanatory variable in the 2004 data set. Generally, flights with layovers (and possible airline changes) incur cost premiums over and above their non-stop counterparts. However, in this particular instance, the variable had a negative coefficient (albeit rather statistically insignificant).

For long-haul flights, Baltimore continues to provide bargains for price-conscious travelers since nearly all of the coefficient values for the remaining airports are positive.

We note that, as with the revised model, Philadelphia's fares experienced a significant shift within this two-year data set. For 2003, its coefficient value was rather high. However, in 2004, its value (-23.45) was even more negative than its corresponding quantity in the revised model. In fact, it nearly approached some degree of statistical significance with a p-value of 0.1089. From 2003 to 2004, State College's fares became far less attractive as its coefficient value was the highest of any of the airports in the latter year.

The consideration of short-haul flights leads to a different set of conclusions. Even though Baltimore still offers bargains (coefficient values for all remaining airports are positive in both 2003 and 2004), Harrisburg becomes an attractive city for airline fares. In 2003, its coefficient value of 23.35 was the lowest of the five remaining airports (excluding Baltimore) while in 2004, its value was the second lowest (only bested by Philadelphia). Perhaps this suggests that Harrisburg may offer competitive fares for short-haul flights, but it fails to provide attractive prices for longer trips.

Unlike the revised and long-haul cases for 2004, Philadelphia's short-haul coefficient was positive (and nearly significant with a p-value of 0.1262). Perhaps Philadelphia's competitiveness lies in its ability to provide great deals for long-haul flights. Choosing Philadelphia for short-haul trips may not offer the impressive fares one tends to experience on long-haul flights. State College deteriorated in performance from 2003 to 2004, as its coefficient in the latter year was the highest of any of the airports.

Closely examining the regression results, however, indicates that this model may lack explanatory power. Although the coefficient of determination increased from 2003 to 2004, the value is still quite a bit lower than the corresponding values experienced in our other regression cases. Granted, some of the explanatory variables in the short-haul model are highly significant, but others have limited explanatory ability. Perhaps, certain nonlinearities are involved when considering the airfare structure for short-haul flights. Or, there may be other variables that could be incorporated to enhance the explanatory ability of the multiple regression model. Whatever the reason, we would need to discover ways of improving this case of the regression model before having increased confidence in its findings.

4. Concluding Remarks

We have developed a multiple regression model to explore the pricing of airline fares for several airports in the U.S. northeast. By analyzing fares for numerous trips within a two-year time frame, as well as flight mileage and number of layovers, we were able to determine those variables that significantly explain airfares.

Southwest's entry into Philadelphia in 2004 may have played a role in that airport's enhanced competitiveness. It had many more instances of least-cost fares in 2004 than it had in 2003. Moreover, its coefficient values from our regression model further confirm that it became a great bargain (even surpassing the deals received in Baltimore). We note that its attractiveness may consist in long-haul flights, as its performance somewhat weakened with short-haul trips.

Harrisburg seems well suited for short-haul flights (although nonlinear models, or other variables, may be required to more fully explain the price structure of these flights). Due to its price premiums over and above competing airports, Harrisburg is definitely not the best choice for long-haul flights.

State College provided fares that were reasonably competitive in 2003 but became less so in the following year. Whether this is a single occurrence or part of a general trend cannot be fully known without conducting similar analyses in subsequent years. Still, it does seem to show a certain volatility in fares for the airports within our sample.

For those of us who teach statistical methods, we face the unenviable task of encountering relatively few students who inherently find our material especially captivating or interesting. That said, we have used this data set as a case study within an upper-year undergraduate decision sciences course and found the response to be quite encouraging. Students appear to be motivated by particular topics that appeal to them. Given that many of our students are rather familiar with airline operations, the analysis of airfares has proven to be a worthwhile pedagogical exercise in our classes.

REFERENCES

Lin, J-S., Dresner, M., and Windle, R. (2001), "Determinants of Price Reactions to Entry in the U.S. Airline Industry," *Transportation Journal* 41 (2/3), 5-22.

U.S. Department of Transportation (1999). "Competition in the U.S. Domestic Airline Industry: The Need for a Policy to Prevent Unfair Practices," Washington, DC.

U.S. Department of Transportation (USDOT), Bureau of Transportation Statistics, Office of Airline Information (2001). *Airport Activity Statistics of Certified Route Air Carriers*. Washington, DC: annual issues, tables 2, 3, 4 and 5.

Windle, R. and Dresner, M. (1995). "Airport Choice in Multiple-airport Regions," *Journal of Transportation Engineering* 121 (4), 332-337.

Appendix

50 Most Populous Cities in the U.S.[1]

1. New York
2. Los Angeles
3. Chicago
4. Houston
5. Philadelphia
6. Phoenix
7. San Diego
8. Dallas
9. San Antonio
10. Detroit
11. San Jose
12. Indianapolis
13. San Francisco
14. Jacksonville
15. Columbus
16. Austin
17. Baltimore
18. Memphis
19. Milwaukee
20. Boston
21. Washington, D.C.
22. Nashville
23. El Paso
24. Seattle
25. Denver
26. Charlotte
27. Portland, OR
28. Oklahoma City
29. Tucson
30. New Orleans
31. Las Vegas
32. Cleveland
33. Long Beach
34. Albuquerque
35. Kansas City
36. Fresno
37. Virginia Beach[2]
38. Atlanta
39. Sacramento
40. Oakland
41. Tulsa
42. Omaha
43. Minneapolis
44. Honolulu
45. Miami
46. Colorado Springs
47. St. Louis
48. Wichita
49. Santa Ana
50. Pittsburgh

[1] The cities of Fort Worth (#27) and Mesa (#42) were excluded due to their proximity to other more populous cities in this list. Fort Worth is served by the Dallas airport, while Mesa is near to Phoenix.
[2] Virginia Beach is served by the Norfolk International Airport.