Empirics of Airline Pricing♦

[Think about a interesting title that will motivate people to read your paper]

[you can use this file as a template for your paper. The letters in green are comments and the letters in black are part of the paper.]

Jaime Palillo♣

[Don't forget your name and your affiliation on the footnote]

September 14, 2016

[Date is important too!]

Abstract

This paper uses online data to analyze the dynamics of airline ticket prices as the flight date nears. We test whether the purchase date affects the price of the ticket. [The abstract should be about 100 words long. It should mention your data source, your hypothesis and your main results. Tell the reader why your paper is important.]

Keywords: Pricing; Advance Purchases; Airlines. [Think about what keyword someone would need to type in Google to find your paper.]

JEL classification: C23; L11; L93. [Pick the two or three codes from the list at http://www.aeaweb.org/journal/jel_class_system.html that identify your area of study.]

1. Introduction [about half a page]

This paper uses an original dataset about airline ticket prices collected from the online travel agency expedia.com. This paper is the first to look at the dynamics of fares as the

♦ I thank my Econ 3341 students at UTRGV for giving me to motivation to write this paper.

♣ Department of Economics & Finance, The University of Texas Rio Grande Valley, 1201 W University Drive, Edinburg, TX 78539. E-mail: jpalillo@utpa.edu

flight date approaches and to distinguish between marginal cost pricing and price discrimination.

Earlier work on price dispersion in airlines can be divided in two. The theoretical work presented in papers like Dana (1999), Prescott (1975) and Eden (1990), and the empirical work presented in Stavins (2001). Dana (1999) presents a model of price dispersion for homogeneous goods, where different prices can be explained by the combination of demand uncertainty and costly capacity.

[Motivate the reader to keep reading your work! Explain why this topic is important. Quick summary of other papers/literature related to your work.]

2. The Data

[Explain your data here. What variables do you have?]

[Provide a summary statistics table]

Table 1: Summary Statistics

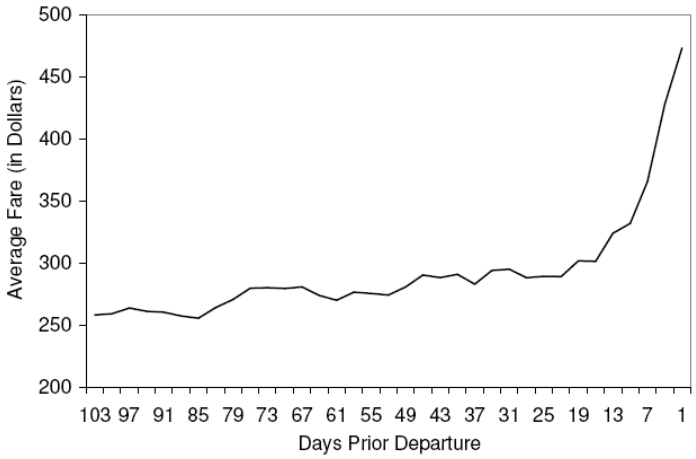
	Mean	Standard Deviation	Minimum	Maximum	Observations
FARE (US$)	291.087	171.879	54.000	1224.000	7933
DAYADV	52.289	30.154	1.000	103.000	7933
LOAD	0.882	1.172	0.022	1.000	7933
DIST	1104.380	620.720	91.000	2604.000	7933
ROUSHASEA	.665	.314	.119	1.000	7933
HHI	.684	.287	.259	1.000	7933
DIFRAIN	2.010	1.484	.000	4.900	7933
AVEHHINC (US$)	35580	4620	25198	53430	7933
AVEPOP	1044072	631862	187704	2897818	7933

[Show any graph that may be interesting to see. (optional)]

Figure 1 shows average fares prior to departure.

Figure 1

Average Fares at Different Days from Departure



[Make sure you explain your graphs and charts/tables.]

Figure 1 shows that as the departure date nears, fares are higher.

3. Econometric Model

[Explain your econometric model here. Write down the equation you will estimate.]

The basic model that we will estimate is:

$$\ln FARE_i = \beta_0 + \beta_1 LOAD_i + \beta_2 DAYADV_i + \beta_3 HHI*DAYADV_i + \beta_4 DIST_i + \beta_5 DISTSQ_i + \beta_6 ROUSHARE_i + \beta_7 HHI_i + \beta_8 DIFRAIN_i + \beta_9 AVEHHINC_i + \beta_{10} AMEANPOP_i + u_i \quad (2)$$

The dependent variable is the logarithm of the price of the ticket and the main coefficient of interest is β_2, the effect of days in advance on the logarithm of prices. A detailed description of the variables appears on Appendix A.

[You can provide a short description of the variables in this section or in the appending (see appendix A)]

4. Results

[Show the regression results. Explain size and magnitude of the coefficient. Are they statistically significant? The standard way to show regression results is:]

Table 2: Estimation Results

	(1)		(2)	
Variables	Coefficient	t-statistic	Coefficient	t-statistic
LOAD	.092	(13.470)	.163	(8.868)
DAYADV	-.003	(-12.395)	-.003	(12.198)
HHI DAYADV*			-.091	(-4.388)
DIST	.002	(37.285)	.002	(37.180)
DISTSQ	-3.4e-7	(-25.577)	-3.4e-7	(-25.435)
ROUSHARE	.252	(5.818)	.254	(5.866)
HHI	-.079	(-1.660)	.066	(1.119)
DIFRAIN	-.0171	(-33.264)	-.174	(-33.305)
AVEHHINC	1.7e-5	(12.562)	1.7e-5	(12.515)
AVEPOP	-1.2e-7	(-11.844)	-1.2e-7	(-11.554)
R-square	.482		.484	

The independent variable is log(*FARE*), *N* = 7933 with 228 routes. *t*-statistics (in parenthesis) are based on heteroscedasticity robust standard errors.

[In Table 2 I am only showing the results for two specifications. One specification does not include the interaction term *HHI*DAYADV* and the other one includes it. You are encouraged to show more specifications, let's say: *FARE* in levels rather than in logarithm, another one without distance squared. Show more than one table here if you want or maybe place additional tables in the appendix.]

The results show that the coefficient on *DAYADV* is negative and significant (the t-statistics is lower than the critical value at any significance level). For every day that you delay in buying the ticket, the price will increase by 0.3 percent.

[Explain your results in detail.]

[Here is where you want to show what you learned in this econometrics class. Some ideas are:

(i) Tests of Hypothesis on the coefficients:

a. t-tests on single slope coefficients.

b. F-test on all the slope parameters.

c. F-test on a subset of slope parameters.

(ii) Confidence intervals.
(iii) Transformations of variables (logs)
(iv) Dummy variables.
(v) Nonlinearities (squared terms)
(vi) Interactions.
(vii) Multicollinearity test
(viii) Heteroscedasticity test
(ix) Specification bias. Are you missing any important variable?
(x) Binary choice models.

The key point here is that you have to show me that you know what you are doing. Discussing your findings and think about your data and the problem in hand. Do not type commands in Gretl without knowing the meaning. I want to see the interpretation applied to the specific problem you want to address. Keep always in mind your main hypothesis. You do not need to write a long paper to get full score. You can base your analysis on examples we did in class.]

[Use the appendices if you want to print out your Gretl output. Make sure to keep your paper between 4 and 6 pages long (without including appendices). Double spacing and font size 12.]

5. Conclusions

Summarize your main results.

References

[Make sure you provide full reference of all the papers/books/other that you cited in the main text.]

Dana, Jr., J. (1999a): "Using yield management to shift demand when the peak time is unknown," *RAND Journal of Economics*, Vol. 30, No. 3, pp. 456-474.

Eden, Benjamin. 1990. "Marginal Cost Pricing When Spot Markets are Complete." *Journal of Political Economy.* Vol. 98, No. 4, pp. 1293-1306.

Prescott, Edward C. 1975. "Efficiency of the Natural Rate." *Journal of Political Economy* Vol. 83 No. 3, pp. 1229-1236.

Stavins, J. (2001): "Price discrimination in the airline market: the effect of market concentration," *The Review of Economics and Statistics*, pp 200 – 202.

Appendix A. Variable description.

FARE: Price in US$ paid for the one-way airfare.

DAYADV: Number of days in advance the ticket was purchased.

LOAD: Ratio of occupied seats to total seats in the aircraft at the moment of purchase.

DIST: Nonstop mileage between the two endpoint airports on a route.

DISTSQ: Distance square.

ROUSHARE: Carrier's share on the route, considering only the direct flights for the day of the flight.

HHI: Herfindahl-Hirshman Index of concentration on the observed route, with *ROUSHARE* used as the measure of market share of each carrier.

DIFRAIN: Absolute difference in average end of October precipitation, measured in inches, between the departure and destination cities.

AVEHHINC: Average of the median household income in the two cities.

AVEPOP: Average population in the two cities. For cities with more than one airport, the population is apportioned to each airport according to each airport's share of total enplanements. Source: Table 3, Bureau Transportation Statistics, Airport Activity Statistics of Certified Air Carriers: Summary Tables 2000.