Price Discrimination through Refund Contracts in Airlines

Diego Escobari[†] Paan Jindapon[‡]

[†] Department of Economics and Finance
The University of Texas - Pan American

[‡] Department of Economics, Finance and Legal Studies
University of Alabama

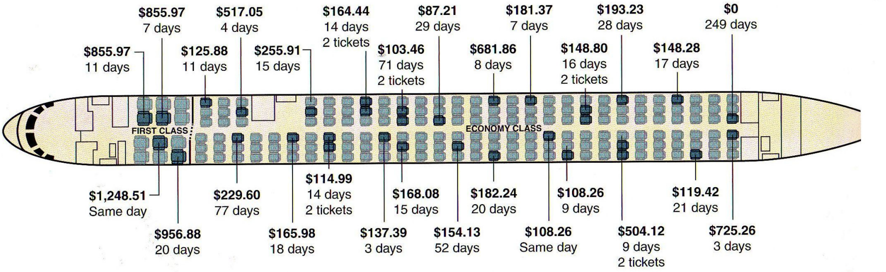
California State University, Fullerton
February, 2012

Outline

1. Introduction
   - Motivation
   - Contribution and intuition of the current paper
2. Theoretical Analysis
   - The Consumer's Problem
   - The Airline's Problem
   - Equilibrium Prices
   - Empirical Implications
3. Empirical Analysis
   - Data
   - Empirical Model
   - Results
4. Conclusions

Motivation: Price dispersion in airlines

Figure: Price dispersion in airlines



- 33 passengers paid 27 different fares, United flight from Chicago to Los Angeles *(New York Times)*
- Gerardi and Shapiro (JPE, 2009)
  Borenstein and Rose (JPE, 1994): 36% difference.

Motivation: Dynamic pricing in airlines

- Key characteristics:
    - Fixed capacity.
    - Perishable good.
    - Aggregate demand uncertainty.
    - Advance sales.
- Carriers exploit ‘fences’ such as:
    - Saturday-night-stayover.
    - Advance purchase discounts.
    - Minimum- and maximum-stay.
    - Refundable tickets.
    - Frequent flier miles.
    - Blackouts.
    - Volume discounts.
    - Fare classes (e.g. coach, first class)
- Airlines have the most sophisticated pricing systems in the world.

Contribution and intuition of the current paper

- Explains how a seller offers refundable/non-refundable tickets in advance to differentiate buyers.
- Can include risk averse consumers [Courty and Li (REStud, 2000), Akan *et at.* (2008) only risk neutral].
- The difference in fares = refundability value + price discrimination.
- First empirical paper in airlines that perfectly controls for observed and unobserved sources of costs.
- First empirical paper that explains the use of non-refundable prices.
- First empirical paper that shows individual demand learning.

The Consumer's Problem

Consumer's type, $i = H$ (high) or L (low), is not observable by the airline.

Period 1:

- Each consumer i decides to buy or not.

Period 2:

- State-dependent utility function:
- State T (Travel):
  Demand $= 1$, with probability π_i.
- State NT (Not Travel):
  Demand $= 0$, with probability $1 - \pi_i$.

The Consumer's Problem

- The valuation of traveling is $v_i > 0$.
- u is the utility of traveling, with $u' > 0$, $u'' < 0$, and $u(0) = 0$.
- Expected utility from buying a refundable ticket at price p:

$$U_i^r(p) = \pi_i u(v_i - p)$$

- Expected utility from buying a non-refundable ticket at price p

$$U_i^{nr}(p) = \pi_i u(v_i - p) + (1 - \pi_i) u(-p)$$

Utility is zero in both states if not buying any ticket.

The Consumer's Problem

- Type i's reservation price for a non-refundable ticket is c_i, such that $U_i^{nr}(c_i) = 0$; i.e.,

$$\pi_i u(v_i - c_i) + (1 - \pi_i) u(-c_i) = 0$$

Example 1:

- $u(x) = \ln(1 + x/1000)$.
- $v_L = 500$.
- $\pi_L = 0.6$.
- We find that: $c_L = 268$.
- The reservation price for a non-refundable ticket is lower.

The Consumer's Problem

- Let the airline offer the menu (p^{nr}, p^{r}).
- Consumer can buy a refundable ticket, a non-refundable ticket, or not buy any ticket.
- Consumer's best response is illustrated in Figure 1.

The Consumer's Problem

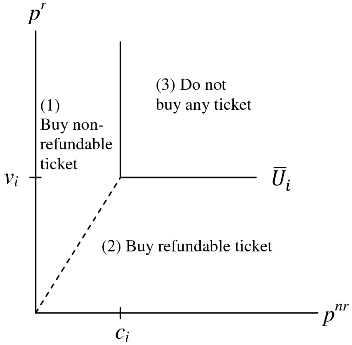


Figure: Type i consumer's best response in (p^{nr}, p^r) space

$$\overline{U}_i \equiv \max\{U_i^{nr}(p^{nr}), U_i^r(p^r)\} = 0$$

The Airline's Problem

- Let the number of type L consumers be N_L. Then $n_L = \pi_L N_L$.
- Let the number of type H consumers be N_H. Then $n_H = \pi_H N_H$.
- The airline announces p^{nr} and p^r at the beginning of period 1.
- Consumers strategies could be either pooling or separating.
- We are interested in a separating equilibria.
- Assume $v_H > v_L$ and $\pi_H < \pi_L$.

The Airline's Problem

- The airline's optimization problem is:

$$\max_{p^{nr},p^{r}} N_L p^{nr} + n_H p^{r}$$

s.t.

$$U_H^r(p^r) \geq U_H^{nr}(p^{nr})$$

$$U_L^{nr}(p^{nr}) \geq U_L^{nr}(p^r)$$

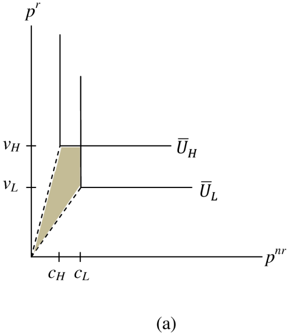
$$U_H^r(p^r) \geq 0$$

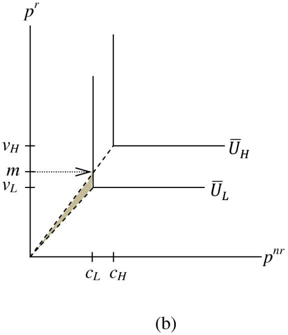
$$U_L^{nr}(p^{nr}) \geq 0.$$

- First two are the incentive-compatibility constraints.
- Last two are the participation constraints. Figure 2 illustrates two cases for the solutions.

The Airline's Problem

Figure: *H* buy refundable tickets and *L* buy non-refundable.





The Airline's Problem

Example 2: Panel (a)

- $u(x) = \ln(1 + x/1000)$.
- $v_L = 500$ and $\pi_L = 0.6$, then we find that: $c_L = 268$.
- $v_H = 800$ and $\pi_H = 0.3$, then we find that: $c_H = 185$.
- Since $c_L \geq c_H$, the airline sets $(p^{nr}, p^r) = (268, 800)$.

Example 3: Panel (b)

- $u(x) = \ln(1 + x/1000)$.
- $v_L = 500$ and $\pi_L = 0.6$, then we find that: $c_L = 268$.
- $v_H = 800$ and $\pi_H = 0.5$, then we find that: $c_H = 185$.
- Since $c_L < c_H$, the airline sets $(p^{nr}, p^r) = (268, 678)$.

The Airline's Problem

Proposition 1 The airline's optimal price menu so that type L consumers buy non-refundable tickets and type H consumers buy refundable tickets is

(a) $(p^{nr}, p^{r}) = (c_L, v_H)$ if $c_L \geq c_H$ or

(b) $(p^{nr}, p^{r}) = (c_L, m)$ if $c_L < c_H$.

Proof See paper.

Equilibrium Prices

- Necessary and sufficient conditions for the airline to find the separating response most profitable.

Proposition 2 Necessary and sufficient conditions for the existence of an equilibrium where the airline sets prices so that type L consumers buy non-refundable tickets and type H consumers buy refundable tickets are

$$\frac{N_H}{N_L} \geq \frac{\pi_L v_L - c_L}{\pi_H (v_H - v_L)}$$

if $c_L \geq c_H$ and

$$\frac{\pi_L v_L - c_L}{\pi_H (m - v_L)} \leq \frac{N_H}{N_L} \leq \frac{c_L}{\pi_H (v_H - m)}$$

if $c_L < c_H$.
Proof See Appendix in the paper.

Empirical Implications

- Let π_H and π_L be dependent on τ (time to departure).
- π_L increases as τ decreases.
- We move from case (a) to case (b).
- The gap between refundable and non-refundable prices diminishes as the flight date nears and the consumers are more certain about their travel plans.
- Only one price prevails at departure ($\tau = 0$)

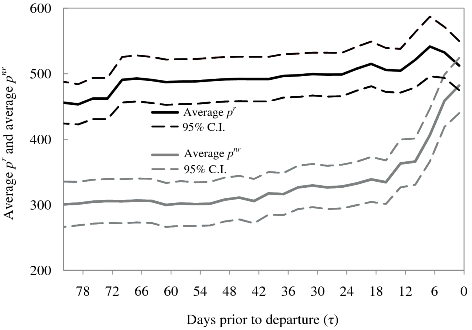
Construction of the Data

- Refundable and non-refundable fares from *expedia.com*
- Pick a single day: Thursday, June 22, 2006.
    - Controls for systematic peak load pricing.
- One-way, non-stop, economy-class..
    - Connecting passengers / sophisticated itineraries / legs.
    - Uncertainty in the return portion of the ticket.
    - Saturday-night-stayover / min- and max-stay.
    - Fare classes (e.g. coach, first class).
- Monopoly routes.
- Panel with 96 cross sectional observations (city pairs).
- Collected every 3 days with 28 observations over time.
- American, Alaska, Continental, Delta, United and US Airways.

Expedia

Data

Figure: Average p^r and p^{nr} with 95% confidence intervals



Controlling for Costs

Costs that change from seat to seat:

- Borenstein and Rose (JPE, 1994)
  - Systematic peak-load pricing.
  - Stochastic peak-load pricing.
- Dana (RAND, 1999)
  - Operational marginal cost.
  - Effective cost of capacity.

Both prices are set for the same seat.

Nonparametric Panel Regression

$$\ln(p^{r}_{ijt} - p^{nr}_{ijt}) = g(\tau_{ijt}, LOAD_{ijt}) + \nu_{ij} + \varepsilon_{ijt}$$

i: flight; j: route; t: time.

Controls for:

- Time-invariant flight-, route-, and carrier-specific characteristics.
  - e.g. systematic peak-load pricing, distance, aircraft type, airport characteristics, managerial capacity.
- Time-variant seat-specific characteristics.
  - e.g. stochastic peak-load pricing, capacity constraints, aggregate demand uncertainty ($\pi_l = \pi_h$).

Nonparametric Panel Regression

$$\ln(p^{r}_{ijt} - p^{nr}_{ijt}) = g(\tau_{ijt}, LOAD_{ijt}) + \nu_{ij} + \varepsilon_{ijt}$$

- $g(\cdot)$: Unknown smooth function.
- Flight-specific effects are outside to avoid the curse of dimensionality.
- Estimated using kernel methods for mixed data types [Racine and Li (J. Econometrics, 2004) and Li and Racine (2007)].
- Better finite sample properties than other kernel estimators.
- Under $\pi_L \neq \pi_H$, include capacity utilization, *LOAD*.
- Allows for interactions among τ and *LOAD* as well as nonlinearities in and among both variables.
- Smoothing parameters will be estimated with least-squared cross-validation.

Dynamic Panels

$$\begin{aligned} \ln(p^{r}_{ijt} - p^{nr}_{ijt}) &= \alpha \ln(p^{r}_{ij,t-1} - p^{nr}_{ij,t-1}) + \beta_1 \tau_{ijt} + \beta_2 \tau^{2}_{ijt} \\ &+ \beta_3 \tau^{3}_{ijt} + \beta_4 LOAD_{ijt} + \nu_{ij} + \varepsilon_{ijt}, \end{aligned} \quad (1)$$

- Nonlinearities in time are modeled parametrically.
- Potential endogeneity of *LOAD*.
    - Estimate using GMM dynamic panels to assume only weak exogeneity of *LOAD*.
    - Rational passengers are allowed to behave dynamically.
    - Controls for potentially serially correlated demand shocks.

Dynamic Panels

- To allow for correlation between carrier effects. (airline specific shocks)
  - Cluster robust standard errors, clustered by airline.
- Difference GMM. Weak instruments when series are persistent.
  - Standard weak instrument test do not work. Use known biases if weak instruments are present.
  - System GMM.
- Moment conditions assume the error term is not serially correlated.
  - Include a second order serial correlation test.
- Test for validity of the instruments.
  - Sargan and Difference Sargan.

Summary Statistics

Table: Summary statistics

Variables	Mean	Std. Dev.	Min.	Max.	Obs.
p^r					
overall	494.486	169.181	144.000	1715.310	2628
between		156.974	144.000	735.497	96
within		64.167	141.262	1474.299	27.375[a]
p^{nr}					
overall	327.749	171.588	64.000	914.000	2628
between		156.654	74.107	665.786	96
within		70.204	164.642	852.249	27.375[a]
τ	41.500	24.238	1.000	82.000	2688
LOAD	0.591	0.241	0.038	1.000	2688

Notes: [a] Number of observations in time, with one observation every three days.

Table: Regression estimates, separate day dummies

	(1)		(2)	
4 days	0.527	(1.575)	0.510	(1.518)
7 days	2.106^{***}	(5.375)	2.067^{***}	(5.361)
10 days	2.614^{***}	(5.290)	2.523^{***}	(5.502)
13 days	2.565^{***}	(4.141)	2.451^{***}	(4.297)
16 days	2.977^{***}	(4.306)	2.800^{***}	(4.613)
19 days	3.013^{***}	(4.190)	2.803^{***}	(4.505)
22 days	2.999^{***}	(5.066)	2.737^{***}	(5.800)
25 days	2.963^{***}	(4.708)	2.674^{***}	(5.423)
28 days	3.036^{***}	(4.751)	2.727^{***}	(5.623)
31 days	3.092^{***}	(4.903)	2.737^{***}	(6.072)
34 days	3.124^{***}	(5.081)	2.733^{***}	(6.471)
37 days	3.187^{***}	(5.081)	2.757^{***}	(6.465)
40 days	3.069^{***}	(4.980)	2.609^{***}	(6.587)
43 days	3.418^{***}	(6.369)	2.927^{***}	(9.515)
46 days	3.325^{***}	(6.522)	2.809^{***}	(10.095)
49 days	3.331^{***}	(6.698)	2.796^{***}	(10.948)
52 days	3.442^{***}	(7.401)	2.878^{***}	(12.559)
55 days	3.439^{***}	(7.441)	2.863^{***}	(12.935)
58 days	3.392^{***}	(7.049)	2.795^{***}	(12.379)
61 days	3.429^{***}	(7.006)	2.818^{***}	(11.960)
64 days	3.291^{***}	(5.722)	2.665^{***}	(8.860)
67 days	3.249^{***}	(5.028)	2.601^{***}	(7.173)
70 days	3.257^{***}	(4.891)	2.600^{***}	(6.902)
73 days	2.993^{***}	(3.390)	2.315^{***}	(3.978)
76 days	3.003^{***}	(3.425)	2.315^{***}	(4.053)
79 days	2.861^{***}	(3.194)	2.161^{***}	(3.686)
82 days	3.177^{***}	(4.309)	2.469^{***}	(5.897)
LOAD			-1.350^{**}	(-2.138)
Within R-squared	0.312		0.319	

Notes: The dependent variable is $\ln(p^{r}_{ijt} - p^{nr}_{ijt})$ and the number of observations is 2628. t-statistics in parentheses based on cluster-robust standard errors, clustered by airline; ***p-value$<$0.01, **p-value$<$0.05, *p-value$<$0.1. Both specifications estimated with flight fixed effects. The 1 day in advance dummy variable excluded.

Table: Regression estimates

	(1) OLS levels	(2) Within groups	(3) GMM *Dif* $t-2$	(4) GMM *Dif* $t-3$	(5) GMM *Sys* $t-2$	(6) GMM *Sys* $t-3$
$\ln(p^{r}_{ij,t-1} - p^{nr}_{ij,t-1})$	0.854^{***}	0.530^{***}	0.572^{***}	0.554^{***}	0.566^{***}	0.560^{***}
	(22.736)	(10.772)	(6.221)	(6.073)	(6.679)	(6.168)
$\tau_{ijt}/10^2$	7.970^{**}	11.578^{**}	10.349^{***}	12.423^{***}	11.782^{***}	12.043^{***}
	(2.222)	(2.537)	(2.610)	(3.860)	(6.475)	(5.411)
$\tau^2_{ijt}/10^4$	-17.228^{*}	-24.928^{*}	-21.435^{***}	-25.486^{***}	-25.525^{***}	-26.080^{***}
	(-1.865)	(-1.891)	(-4.211)	(-4.782)	(-5.019)	(-5.183)
$\tau^3_{ijt}/10^6$	11.236	15.898	13.680^{***}	16.220^{***}	16.783^{***}	17.152^{***}
	(1.608)	(1.501)	(4.077)	(4.389)	(4.136)	(4.353)
$LOAD_{ijt}$	-0.434^{***}	-0.828^{**}	-0.068	0.102	-0.317	-0.289
	(-5.578)	(-2.156)	(-0.026)	(0.064)	(-0.129)	(-0.124)
Serial correlation test^{a}(p-value)			0.605	0.619	0.604	0.609
Sargan test^{b}(p-value)			0.004	0.066	0.689	0.988
Difference Sargan test^{c}(p-value)					1.000	1.000

Notes: The dependent variable is $\ln(p^{r}_{ijt} - p^{nr}_{ijt})$. Columns 2 through 6 control for carrier-, route-, and flight-specific characteristic. t-statistics in parentheses for the OLS and the Within groups based on cluster-robust standard errors, clustered by airline. t-statistics in parentheses for the two-step system GMM based on Windmeijer WC-robust estimator; ***p-value<0.01, **p-value<0.05, *p-value<0.1. [a] The null hypothesis is that the errors in the first-difference regression exhibit no second-order serial correlation (valid specification). [b] The null hypothesis is that the instruments are not correlated with the residuals (valid specification). [c] The null hypothesis is that the additional instruments $t-3$ are not correlated with the residuals (valid specification).

Regression Estimates

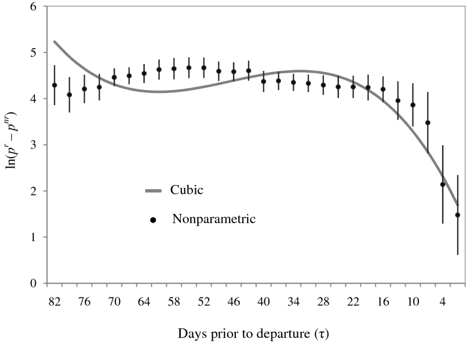
Table: Regression estimates, robustness checks

	(1) Within groups	(2) Within groups	(3) GMM *Sys* $t-2$	(4) GMM *Sys* $t-3$	(5) Within groups	(6) Within groups	(7) GMM *Sys* $t-2$	(8) GMM *Sys* $t-3$	(9) Within groups	(10) Within groups
$\ln(p^{r}_{ij,t-1} - p^{nr}_{ij,t-1})$			0.723	0.717			0.616	0.609		
			(9.706)	(9.796)			(6.948)	(6.813)		
$\tau_{ijt}/10^2$	1.783	-0.084	-0.107	-0.104	9.421	8.200	5.600	5.630	18.210	17.004
	(3.916)	(-0.641)	(-0.233)	(-0.137)	(12.128)	(18.801)	(2.859)	(2.645)	(3.381)	(3.584)
$\tau^2_{ijt}/10^4$					-9.117	-8.506	-5.360	-5.440	-35.358	-35.111
					(-16.462)	(-14.991)	(-5.914)	(-5.703)	(-2.404)	(-2.470)
$\tau^3_{ijt}/10^6$									20.952	21.295
									(1.791)	(1.842)
$LOAD_{ijt}$		-2.751	-1.633	-1.662		-1.044	0.239	0.162		-1.155
		(-3.961)	(-1.988)	(-1.367)		(-1.734)	(0.102)	(0.062)		(-2.062)
Serial correlation test^a(p-value)			0.503	0.504			0.577	0.581		
Sargan test^b(p-value)			0.676	0.989			0.699	0.989		
Difference Sargan test^c(p-value)			1.000	1.000			1.000	1.000		
Within R-squared	0.105	0.139			0.232	0.237			0.263	0.268

Notes: The dependent variable is $\ln(p^{r}_{ijt} - p^{nr}_{ijt})$. All specifications control for carrier-, route-, and flight-specific characteristic. See notes on Table 5.

Nonparametric Estimation

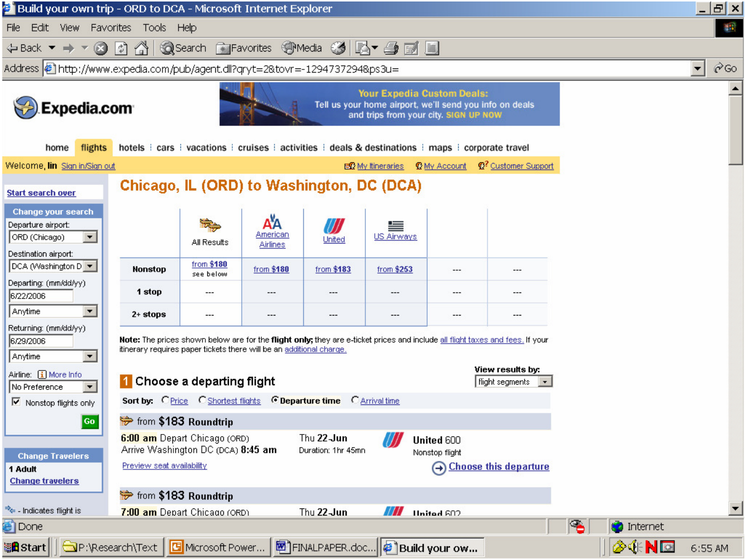
Figure: Nonparametric partial regression plot and cubic specification



- Bivariate plot is holding *LOAD* in its median.

Conclusions

- Importance of offering a menu of prices.
- A seller can price discriminate when heterogeneous buyers are uncertain about their demand for travel.
- Buyers can use refund contracts to insure against uncertainty in consumption.
- The gap between fares is a function of individual's demand uncertainty.
- Nonparametric regression shows that most of the individual demand uncertainty is resolved during the last two weeks.
- The opportunity to price discriminate decreases closer to departure.



Departure airport: ORD (Chicago)
Destination airport: DCA (Washington D
Departing: (mm/dd/yy) 6/22/2006
Anytime
Returning: (mm/dd/yy) 6/29/2006
Anytime
Airline: More Info
No Preference
☑ Nonstop flights only

1 Adult

- Indicates flight is

Chicago, IL (ORD) to Washington, DC (DCA)

	All Results	American Airlines	United	US Airways		
Nonstop	from $180 see below	from $180	from $183	from $253	---	---
1 stop	---	---	---	---	---	---
2+ stops	---	---	---	---	---	---

Note: The prices shown below are for the **flight only;** they are e-ticket prices and include all flight taxes and fees. If your itinerary requires paper tickets there will be an additional charge.

View results by: flight segments

1 Choose a departing flight

Sort by: ○ Price ○ Shortest flights ● **Departure time** ○ Arrival time

from **$183 Roundtrip**

6:00 am Depart Chicago (ORD) Thu **22-Jun** **United** 600
Arrive Washington DC (DCA) **8:45 am** Duration: 1hr 45mn Nonstop flight
Preview seat availability
Choose this departure

from **$183 Roundtrip**

7:00 am Depart Chicago (ORD) Thu **22-Jun** **United** 602

Data