Screening and Price Discrimination with Unobserved Consumer Types

Diego Escobari[†] Manuel A. Hernandez[‡]

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

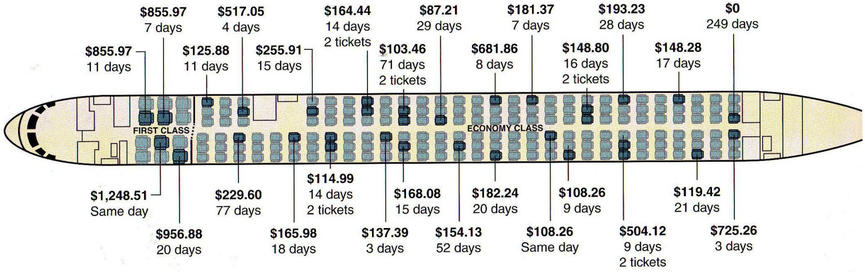
[‡]Markets, Trade and Institutions
International Food Policy Research Institute

Latin American and Caribbean Economic Association
Santa Cruz, Bolivia
October 2015

Outline

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)*
- Borenstein and Rose (JPE, 1994): 36% difference.
- Gerardi and Shapiro (JPE 2009).

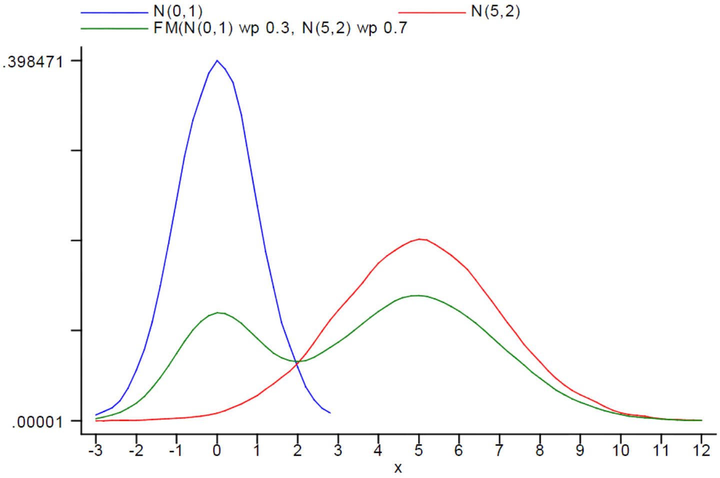
Motivation: Price discrimination in Airlines

- 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)
    - Hour-of-day purchase.
- Airlines have the most sophisticated pricing systems in the world.

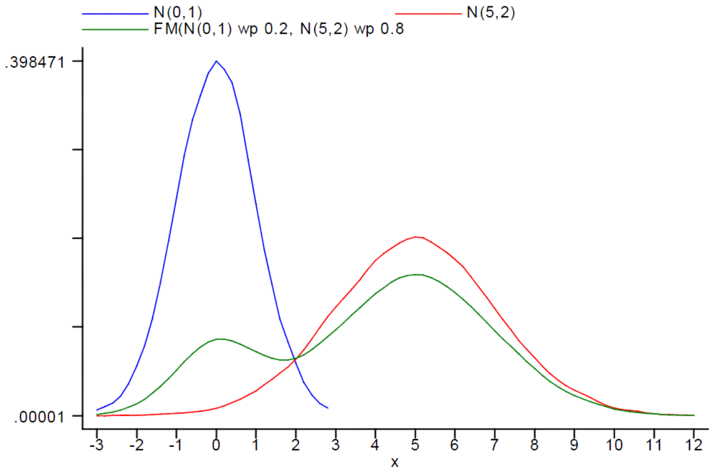
Motivation: Asymmetric Information

- Consumers hold private information (types are unknown to the seller).
- Mechanism design makes buyers reveal information:
  - Differentiated products (menu of prices)
  - Quantity discounts
- Propose using incomplete information to identify unobserved consumer types:
  - Consumers have unit demands.
  - Product is homogeneous.
- Well suited for airlines:
  - No arbitrage opportunities.
  - Price dispersion and consumer heterogeneity.
  - Data on a large number of markets.

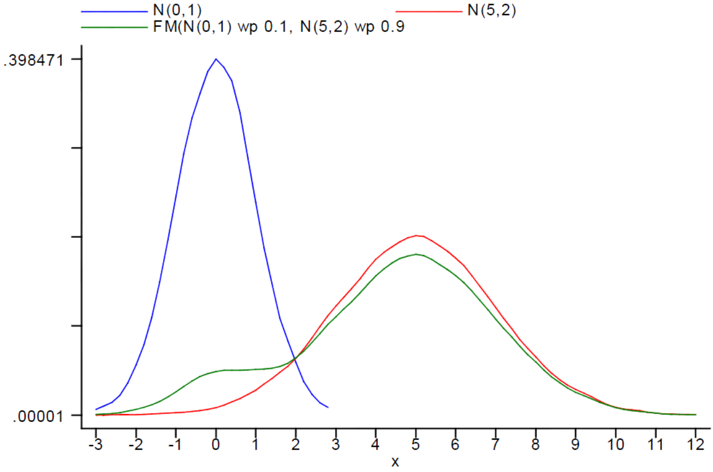
Mixtures: Graphical View



Mixtures: Graphical View



Mixtures: Graphical View



Contribution and Intuition

- Use mixtures to identify consumer types.
- High types (business travelers) have:
  - Less price-sensitive demand.
  - Have higher valuations.
  - Pay higher prices.
- Larger within-type sales dispersion in low types (greater consumer heterogeneity)
- Probability of high-types increases with:
  - Higher capacity utilization.
  - Closer to departure (when fares are low).
  - Income.
  - At hub airports.
  - Market concentration.

Construction of the Data

- Posted prices 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).
- Panel with 228 cross sectional observations (city pairs).
- Collected every 3 days with 35 observations in time.
- American, Alaska, Continental, Delta, United and US Airways.

Expedia

Summary Statistics

Table: Summary Statistics

VARIABLE	mean	sd	min	max
Monopolies:				
FARE (P)	322.6	171.7	64	914
DAYS	50.8	29.3	1	100
LOAD	0.544	0.245	0.013	1
SALES (Q)	0.017	0.038	-0.392	0.485
$\mathrm{I}_{\mathrm{FARE}>\overline{\mathrm{FARE}}}$	0.349	0.477	0	1
$\mathrm{I}_{\mathrm{LOAD}>\overline{\mathrm{LOAD}}}$	0.427	0.495	0	1
Full sample:				
FARE (P)	292.2	172.3	54	1,224
DAYS	50.8	29.3	1	100
LOAD	0.513	0.250	0.013	1
SALES (Q)	0.017	0.042	-0.408	0.485
INCOME	35,580.0	4619.4	25,198	53,430.0
LEISURE	0.070	0.256	0	1
SLOT	0.298	0.458	0	1
HUB	0.737	0.440	0	1
DISTANCE	1104.4	620.7	91	2,604
HHI	0.679	0.289	0.253	1
$\mathrm{I}_{\mathrm{FARE}>\overline{\mathrm{FARE}}}$	0.340	0.474	0	1
$\mathrm{I}_{\mathrm{LOAD}>\overline{\mathrm{LOAD}}}$	0.416	0.493	0	1

Note: The number of observations is 3,243 for the monopolies and 7,705 for the full sample.

Average and standard deviation of fares

- Prices as the flight date nears

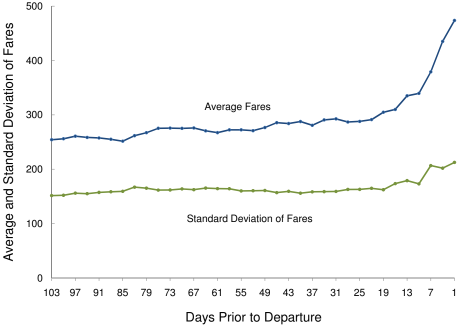


Figure: Average and standard deviation of fares

Demand Equations

N different consumer types:

$$Q_{ijt} = \begin{cases} \alpha_1 + \beta_1 P_{ijt} + X\delta_1 + \kappa_{i,1} + \varepsilon_{ijt,1} & \text{if } \theta = 1, \\ \alpha_2 + \beta_2 P_{ijt} + X\delta_2 + \kappa_{i,2} + \varepsilon_{ijt,2} & \text{if } \theta = 2, \\ \vdots & \vdots \\ \alpha_N + \beta_N P_{ijt} + X\delta_N + \kappa_{i,N} + \varepsilon_{ijt,N} & \text{if } \theta = N, \end{cases}$$

- i: flight; j: route; t: time.
- Unobserved types: $\theta = 1, \ldots, N$
- Q: Sales.
- P: Posted price.
- X: Other factors (days in advance).

Maximum Likelihood

- $\varepsilon_{ijt,\theta} \sim N(0, \sigma^2_{\varepsilon,\theta})$, $\theta = 1, ..., N$, the log-likelihood for the kth flight-time period is:

$$\ln l_k = \ln \left[\sum_{\theta=1}^{N} \frac{r_\theta}{\sigma_{\varepsilon,\theta}\sqrt{2\pi}} \exp\left(\frac{-\varepsilon^2_{k,\theta}}{2\sigma^2_{\varepsilon,\theta}} \right) \right]$$

- where r_θ is the mixing parameter defined as the probability of being in a regime dominated by type θ consumers.
- $\sum_{\theta=1}^{N} r_\theta = 1$

Maximum Likelihood

- Each kth observation can be associated to a particular demand regime θ, $\theta = 1, ..., N$, with probability r_θ.
- We can model the probability of being in a type-θ demand as,

$$r_\theta = \frac{\exp\left(G\delta_\theta\right)}{1 + \sum_{s=1}^{N-1} \exp\left(G\delta_s\right)}$$

- G: Observables that can help us identify the type.

Maximum Likelihood Estimates: Price Discrimination

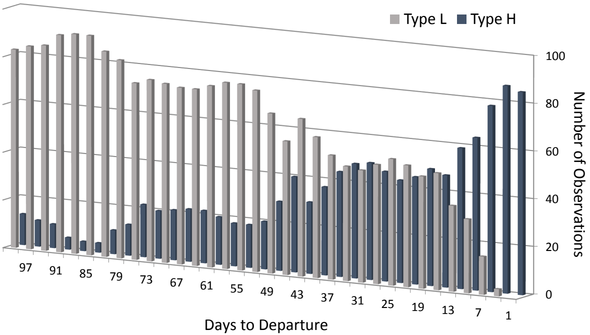
Table: Maximum Likelihood Estimates: Price Discrimination

Model:	(1)	(2)	
Type θ:	Pooled	H	L
Demand Equations:			
CONSTANT	1.6840*	0.0737	3.7413*
	(0.1496)	(0.0955)	(0.3788)
LNFARE	-2.3664*	-1.4578*	-2.6381*
	(0.3409)	(0.1581)	(0.9031)
DAYS	-0.0332*	-0.0175*	-0.0491*
	(0.0027)	(0.0014)	(0.0065)
σ_ε	3.7172*	1.1275*	5.5912*
	(0.0485)	(0.0380)	(0.1361)
Probability of Type H, $r_H = \text{Prob}(\theta = H)$:			
$\mathrm{I}_{\text{FARE} > \overline{\text{FARE}}}$		0.7376*	
		(0.1138)	
Average Fare	322.6	381.6	290.9
Observations	3,243	3,243	
Log likelihood	6,075.1	7,106.9	
SBIC[a]	-3.737	-4.360	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Probability of type-H demand

Figure: Number of observations by type and days in advance



Determining the Number of Types N and Recovering Valuations

Determining the Number of Types N

Table: Determining the Number of Types N

Number of types N	SBIC[a]	Log likelihood	LR test[b]
1	-3.737	6075.057	
2	-4.360	7106.864	0.000
3	-4.351	7111.356	0.110
4	-4.340	7114.844	0.222
5	-4.331	7119.636	0.088

Note: [a] Schwarz Bayesian Information criterion. [b] p-value of likelihood ratio (LR) test reported.

Recovering Valuations

- Reservation values are uniformly distributed $[0, \bar{v}_\theta]$.
- Demand is: $Q = N_\theta - N_\theta / \bar{v}_\theta P$.
- The number of consumers of each type is $N_\theta = \alpha_\theta + X\delta_\theta$.

$$\bar{v}_\theta = -\frac{\alpha_\theta + X\delta_\theta}{\beta_\theta}.$$

Maximum Likelihood Estimates: Reservation Values

Table: Maximum Likelihood Estimates: Reservation Values

Model:	(1)	(2)	
Type θ:	Pooled	*H*	*L*
Demand Equations:			
CONSTANT	1.6529*	0.0787	3.6683*
	(0.1425)	(0.0917)	(0.3607)
FARE	-0.0074*	-0.0046*	-0.0076*
	(0.0010)	(0.0005)	(0.0025)
DAYS	-0.0326*	-0.0174*	-0.0475*
	(0.0025)	(0.0014)	(0.0063)
σ_ε	3.7190*	1.1356*	5.6126*
	(0.0475)	(0.0365)	(0.1326)
Probability of Type H, $r_H = \mathrm{Prob}(\theta = H)$:			
$\mathrm{I}_{\text{FARE} > \overline{\text{FARE}}}$		0.7602*	
		(0.1096)	
Reservation Values:	554.3	692.8	546.4
Average Fare	322.6	381.6	290.9
Observations	3,243	3,243	
Log likelihood	6,073.5	7,106.3	
SBIC[a]	-3.736	-4.360	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Maximum Likehood Estimates: Role of Capacity

Table: Maximum Likelihood Estimates: Role of Capacity

Model:	(1)		(2)	
Type θ:	H	L	H	L
Demand Equations:				
CONSTANT	0.0838	3.4119*	0.1491	3.5532*
	(0.1080)	(0.3692)	(0.1003)	(0.4205)
LNFARE	-1.4496*	-3.4422*	-1.4359*	-2.8008*
	(0.1972)	(0.9093)	(0.1878)	(1.0884)
DAYS	-0.0170*	-0.0431*	-0.0172*	-0.0436*
	(0.0016)	(0.0072)	(0.0015)	(0.0082)
σ_ε	1.1453*	5.7486*	1.2375*	6.0606*
	(0.0523)	(0.1687)	(0.0457)	(0.1901)
Probability of Type H, $r_H = \mathrm{Prob}(\theta = H)$:				
$\mathrm{I}_{\text{FARE} > \overline{\text{FARE}}}$			1.1533*	
			(0.1450)	
$\mathrm{I}_{\text{LOAD} > \overline{\text{LOAD}}}$	0.9974*		0.6630*	
	(0.1930)		(0.1687)	
$\text{DAYS} - \overline{\text{DAYS}}$	0.0261*		0.0303*	
	(0.0031)		(0.0032)	
Observations	3,243		3,243	
Log likelihood	7,137.0		7,180.3	
SBIC[a]	-4.377		-4.401	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Probability of type-H demand

Figure: Probability of type-H demand

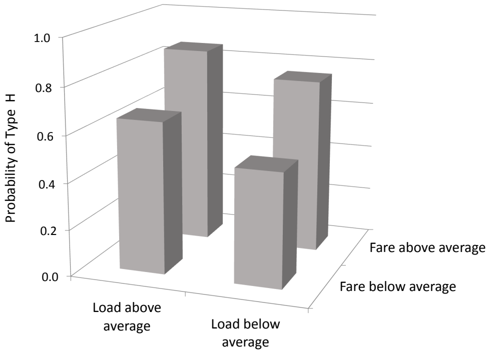


Table: Maximum Likelihood Estimates: Role of Competition

Model:	(1)	(2)		(3)	
Type θ:	Pooled	*H*	*L*	*H*	*L*
Demand Equations:					
CONSTANT	1.8369*	0.0322	3.8670*	0.0632	3.5123*
	(0.1071)	(0.0664)	(0.2842)	(0.0733)	(0.2773)
LNFARE	-1.7219*	-1.0358*	-2.0037*	-0.9848*	-2.0272*
	(0.2048)	(0.1055)	(0.6350)	(0.1136)	(0.6096)
DAYS	-0.0362*	-0.0177*	-0.0522*	-0.0170*	-0.0443*
	(0.0018)	(0.0010)	(0.0050)	(0.0010)	(0.0057)
σ_ε	4.0316*	1.1038*	6.0516*	1.1661*	6.4194*
	(0.0343)	(0.0267)	(0.1003)	(0.0308)	(0.1381)
Probability of Type *H*, $r_H = \text{Prob}(\theta = H)$:					
$\mathrm{I}_{\text{FARE}>\overline{\text{FARE}}}$		0.5266*		0.9872*	
		(0.0750)		(0.0836)	
$\mathrm{I}_{\text{LOAD}>\overline{\text{LOAD}}}$				0.5496*	
				(0.1112)	
$\text{DAYS} - \overline{\text{DAYS}}$				0.0327*	
				(0.0020)	
Observations	7,705	7,705		7,705	
Log likelihood	13,807.7	16,471.5		16,713.5	
SBIC[a]	-3.579	-4.265		-4.326	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Role of Route Characteristics

Table: Maximum Likelihood Estimates: Role of Route Characteristics

Model:	(1)		(2)		(3)	
Type θ:	H	L	H	L	H	L
Demand Equations:						
CONSTANT	0.7170*	5.9771*	0.7208*	6.0161*	0.7209*	6.0107*
	(0.0670)	(0.5449)	(0.0712)	(0.6588)	(0.0680)	(0.6936)
LNFARE	-1.3195*	-2.8756*	-1.3209*	-2.8989*	-1.3235*	-2.8985†
	(0.1133)	(1.0855)	(0.1132)	(1.0175)	(0.1139)	(1.1793)
DAYS	-0.0239*	-0.0749*	-0.0239*	-0.0754*	-0.0239*	-0.0754*
	(0.0011)	(0.0103)	(0.0012)	(0.0109)	(0.0011)	(0.0124)
σ_ε	1.6183*	8.6465*	1.6228*	8.6719*	1.6220*	8.6680*
	(0.0343)	(0.2396)	(0.0325)	(0.2531)	(0.0319)	(0.2706)
Probability of Type H, $r_H = \mathrm{Prob}(\theta = H)$:						
LNINCOME	0.1521*		0.1354*		0.1253†	
	(0.0062)		(0.0096)		(0.0589)	
LEISURE	-0.0695		-0.0004		-0.0855	
	(0.1742)		(0.1649)		(0.2126)	
SLOT			-0.0447		0.0569	
			(0.0928)		(0.1301)	
HUB			0.2679*		0.2663†	
			(0.0963)		(0.1095)	
LNDISTANCE					-0.0292	
					(0.0839)	
HHI					0.4188†	
					(0.1717)	
Observations	7,705		7,705		7,705	
Log likelihood	16,875.4		16,880.2		16,883.8	
SBIC[a]	-4.369		-4.368		-4.366	

Note: The dependent variable is SALES $\times$ 100. Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%. [a] Schwarz Bayesian Information criterion.

Potential Endogeneity of Fares

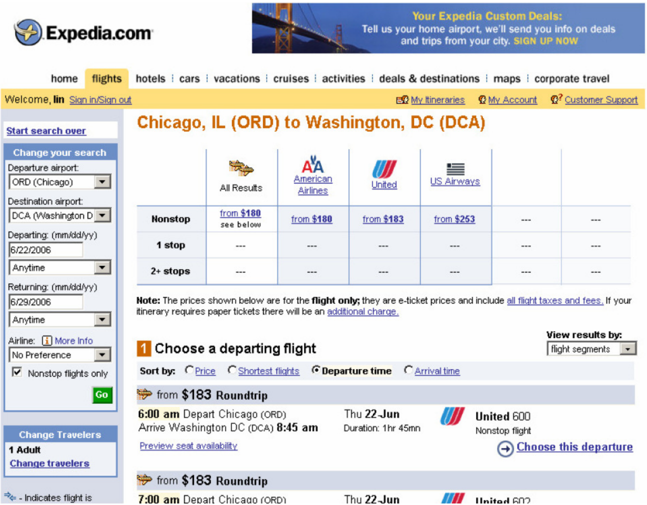
Table: Hausman Test for Potential Endogeneity of Fares

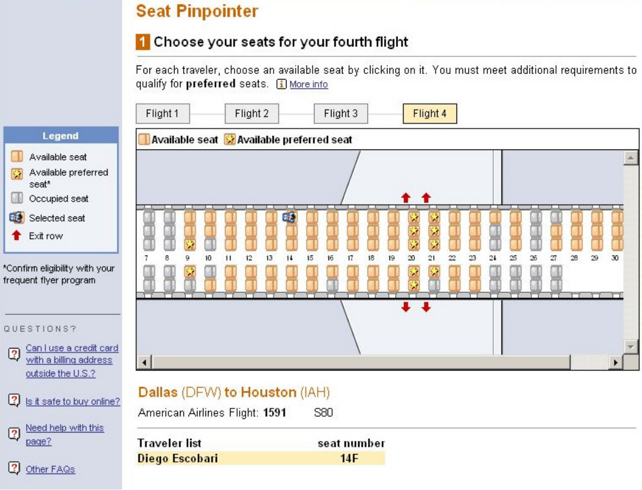
Dependent variable:	LnFare
First Stage Regressions:	
Lag LnFare	-0.0329*
	(0.0044)
Days	-0.0008*
	(0.0001)
Constant	0.2413*
	(0.0279)
Observations	3,145
Underidentification test:	
Kleibergen-Paap rk LM statistic	53.189
$\chi^2(1)$ P-val	0.000
Weak identification test:	
Kleibergen-Paap rk Wald F statistic	55.282
Hausman test. H_0: Fare is exogenous	
F(1,3141)	0.349
Prob $>$ F(1,3141)	0.559

Note: Standard errors in parentheses. All regressions control for flight fixed effects. ‡ significant at 10%; † significant at 5%; * significant at 1%.

Conclusion

- Consumers hold information that is unknown to the seller.
- Present mixtures to separate consumer types.
- Does not rely on particular product attributes as a screening device
- Use partial information: capacity utilization and days to departure.
- Evidence of two types of consumers: Low types more closely resemble “tourists” and high types are business travelers.
- We find that high types are less price sensitive, have higher valuations and pay higher prices.
- The proportion of high types increases as the departure date nears.
- High types are more likely to make a purchase when most travelers already booked.


Expedia.com
Your Expedia Custom Deals:
Tell us your home airport, we'll send you info on deals and trips from your city. SIGN UP NOW
home | flights | hotels | cars | vacations | cruises | activities | deals & destinations | maps | corporate travel
Welcome, lin Sign in/Sign out
My Itineraries
My Account
Customer Support
Start search over
Change your search
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
Go
Change Travelers
1 Adult
Change travelers
- 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.
1 Choose a departing flight
View results by:
flight segments
Sort by: Price Shortest flights Departure time Arrival time
from $183 Roundtrip
6:00 am Depart Chicago (ORD)
Arrive Washington DC (DCA) 8:45 am
Thu 22-Jun
Duration: 1hr 45mn
United 600
Nonstop flight
Preview seat availability
Choose this departure
from $183 Roundtrip
7:00 am Depart Chicago (ORD)
Thu 22-Jun


Seat Pinpointer
1 Choose your seats for your fourth flight
For each traveler, choose an available seat by clicking on it. You must meet additional requirements to qualify for preferred seats. More info
Flight 1
Flight 2
Flight 3
Flight 4
Available seat
Available preferred seat
Legend
Available seat
Available preferred seat*
Occupied seat
Selected seat
Exit row
*Confirm eligibility with your frequent flyer program
QUESTIONS?
Can I use a credit card with a billing address outside the U.S.?
Is it safe to buy online?
Need help with this page?
Other FAQs
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
Dallas (DFW) to Houston (IAH)
American Airlines Flight: 1591 S80
Traveler list
seat number
Diego Escobari
14F

Data

Table: Hausman Test to Evaluate the Model Identification

Excluded variables from type equation	H_0: Difference in coefficients of demand equation between benchmark model and alternative models not systematic
$\mathrm{I}_{\textsc{Fare}>\overline{\textsc{Fare}}}$	11.5929 (0.1148)
$\mathrm{I}_{\textsc{Load}>\overline{\textsc{Load}}}$ & $\textsc{Days}-\overline{\textsc{Days}}$	9.4537 (0.2217)

Note: Benchmark model includes $\mathrm{I}_{\textsc{Fare}>\overline{\textsc{Fare}}}$, $\mathrm{I}_{\textsc{Load}>\overline{\textsc{Load}}}$ and $\textsc{Days}-\overline{\textsc{Days}}$ in the type equation. Hausman Chi-squared statistic reported and *p*-value in parentheses.