Demand shifting across flights and airports in a spatial competition model

Escobari, Diego and Lee, Sang-Yeob

The University of Texas - Pan American, Korea Institute of Public Finance

17 April 2012

Demand Shifting across Flights and Airports in a Spatial Competition Model*

Diego Escobari[†] Sang-Yeob Lee[‡]

April 17, 2012

Abstract

This paper investigates the nature of day-to-day competition between flights using a unique panel data set on prices and inventories. We use instrumental variables methods and several spatial autoregressive models (SAR) to estimate price reaction functions. The primary source of product differentiation is departure time. After controlling for flight-specific characteristics and various sources of price dispersion, we find important evidence of demand shifting between competing flights. Most of the shift is being captured by flights scheduled to depart within a 3-hour window. We find no evidence of demand shifting between airports.

Keywords: Spatial Autoregressive Models, Competition, Demand Shifting, Airlines

JEL Classification: C21, D4, L93

*We thank Paan Jindapon, Lung-fei Lee, Rossitza Wooster, and seminar participant at the Southern Economics Association meetings. Two anonymous referees provided insightful comments that helped improve the paper. Stephanie C. Reynolds provided capable research assistance.

†Department of Economics & Finance, The University of Texas - Pan American, Edinburg, TX 78539, Phone: (956) 665-3366, Fax: (956) 665-5020, Email: escobarida@utpa.edu, URL: http://faculty.utpa.edu/escobarida

‡Korea Institute of Public Finance, Seoul, Korea, Phone: +82-2-2186-2257, Email: syeoblee@kipf.re.kr.

1 Introduction

Analyzing the nature of day-to-day competition and pricing in airlines is particularly important due to the existence of aggregate demand uncertainty and capacity constraints. Airlines sell tickets in advance when there is uncertainty about the aggregate demand. Moreover, capacity if fixed and can only be augmented at a relatively high marginal costs. Hence, carriers employ dynamic pricing techniques, commonly known as 'yield management' to try to reduce the costs of aggregate demand uncertainty and capacity costs. The role of competition is important because a specific flight's pricing decisions have to take into account the prices of the competitors, especially in light of how easy it is for buyers to compare fares across different departure times, dates and carriers.

Existing literature in airlines that analyzes the role of competition (e.g., Borenstein and Rose (1994), and Gerardi and Shapiro (2009)) has focused on the analysis of how market structure affects price dispersion. These papers, like most of the empirical work in airlines, use the DB1B data from the *Bureau of Transportation Statistics*. The restriction of the DB1B is that it is aggregate quarterly data that does not contain the contemporaneous posted prices that are key to assess the existing day-to-day competition in airlines. In this paper we use a unique airline data set obtained from the online travel agency Expedia.com that contains information on prices, inventories, and contemporaneous prices of competing flights. Previous research that uses data from online travel agencies includes Stavins (2001), Bilotkach (2006), Escobari (2009), and Alderighi et al. (2011).

The current paper extends the existing literature by modeling the day-to-day competition between flights. We estimate the degree of demand shifting between competing flights taking into account the most important nature of product differentiation, flight departure times. The structure of the data also allows us to identify the degree of competition and demand shifting across New York City area airports. To achieve this we use IV methods and various spatial autoregressive (SAR) models to estimate price reaction functions. We find that there is important demand shifts between competing flights, with nearly the entire demand shift being captured by flights that depart within 3 hours of the scheduled flight. We find no evidence of demand shifts between airports that serve the New York city area for our particular destination, Toronto.

The rest of the paper is structured as follows. Section 2 describes the collection of the data. The estimation methodology is explained in section 3 and the results reported and discusses in section 4. Section 5 concludes.

2 Data

Following Escobari (2009) and Escobari (2012), we collected from the online travel agency Expedia.com the lowest available one-way economy-class posted fares for all the non-stop flights between the three big airports serving the New York City area and the main airport in Toronto.[1] The methodology allows us to keep track of inventories of seats at each posted price by looking at the seat availability map. Given that overbookings are usually a small fraction of the total number of tickets, our measure is assumed to be proportional to bookings. Moreover, tickets obtained through frequent-flyer programs are excluded from the sample. The three airports in the New York City area are Newark Liberty International Airport (EWR), John F. Kennedy International Airport (JFK), and La Guardia Airport (LGA), while the airport in Toronto is the Toronto Pearson International Airport (YYZ). The data set includes all 317 flights that served this route between December 19 and December 24, 2008.[2] The data set is a panel, where each cross-sectional observation corresponds to a flight. Fares and inventory levels were recorded every three days, with the first cut in time corresponding to 43 days prior to departure and the last to 1 day prior to departure. The carriers in the sample are American, Air Canada, Continental, Delta, Lan Chile and United. Table 1 shows a summary of the flights by day and by carrier.

The construction of this data set as well as the selection of this particular city pair has some important advantages. The route — New York City to Toronto — was selected

[1]While focusing on one-way fares seems restrictive, this should not affect the estimation as long as the carriers adjust one-way prices based on the current inventories. Moreover, our results can be easily generalized to round-trip tickets under the standard assumption in airlines where the round-trip price is assumed to be two times the one-way price. See for example Borenstein and Rose (1994, p. 677), and Gerardi and Shapiro (2009, p. 5).

[2]While the week before Christmas is not the typical week for air travel, it is not clear how the results would be different during a more usual week. Our period of study probably has higher occupancy rates and higher fares (see, for example Escobari (2009)). Sellers are potentially scheduling more or larger flights, while buyers may be exerting more effort in their search for lower fares.

because all New York flights arriving in Toronto have to land at the Toronto Pearson Airport. Moreover, having three departing airports in the New York City area will let us analyze the effect of competition between flights that have similar departure times, same destination, but differ on the departing airport. By picking non-stop flights and one-way fares we control for price differences associated with fare restrictions and cost differences associated with round trip tickets (e.g., as Saturday-night-stay, minimum-, and maximum-stay). This also controls for variation in prices related to more sophisticated itineraries (e.g., travelers connecting to different cities). Because airlines price discriminate and offer at the same time and for the same flight different types of tickets, we selected the least expensive fare to control for the existence of a more expensive refundable ticket. Moreover, restricting the analysis to economy-class tickets is important to control for heterogeneity in consumers who may prefer to fly in first-class.

3 Estimated Model

3.1 SAR and Price Reaction Functions

In this section we describe the specifications that characterize the flights' price reaction functions. The first specification captures the demand shifting across different flights within the same airport,

$$\text{PRICE}_{it} = \lambda_1 \mathrm{W} \cdot \text{PRICE}_{it} + \mathrm{X}\beta + \eta_i + \varepsilon_{it}. \tag{1}$$

The second specification captures the demand shifting across different airports while controlling for competition within the same airport,

$$\text{PRICE}_{it} = \lambda_1 \mathrm{W} \cdot \text{PRICE}_{it} + \lambda_2 \mathrm{M} \cdot \text{PRICE}_{it} + \mathrm{X}\beta + \eta_i + \varepsilon_{it}, \tag{2}$$

where PRICE denotes the price for flight i at time t. X is a matrix of control variables that includes linear and higher order polynomials of ADVANCE and SEATS. ADVANCE is the number of days in advance — prior to the departure date — the fares were recorded, while SEATS measures the inventories of seats relative to the aircraft size. It is calculated as the ratio of seats sold to total seats in the aircraft. The matrices W and M are the spatial weights that correspond to the flights departing from the same airport, W, and the flights departing from different airports, M. η_i denotes the unobservable flight specific effect and ε_{it} denotes the remaining disturbance.

The variables in X and different assumptions on the two-way error component will allow to control for various sources of price dispersion in airlines. The time-invariant flight-specific characteristics (η_i) is very flexible because it controls for carrier- and airport-specific characteristics as well. This includes costs at the flight level, aircraft's characteristics, carrier's managerial capacity, route's Herfindahl index based on the number of flights, distance, and hub and network characteristics that are time invariant. These are all time-invariant characteristics in Stavins (2001), Borenstein and Rose (1994), and Gerardi and Shapiro (2009), who used a cross section of tickets and more aggregate data, respectively. ADVANCE and its higher order polynomials are expected to capture advance purchase restrictions and other common nonlinear trends in prices that may include intertemporal price discrimination strategies. Moreover, higher order polynomials on SEATS are expected to control for the path of inventories and sales as predicted by models that take into account aggregate demand uncertainty and costly capacity (e.g., Prescott (1975) and Dana (1999)).

The degree of demand shifting across flights is captured by the Spatial Autoregressive Coefficient λ_1 in Equation 1. If the price of a flight does not respond to the prices of the competitors, this coefficient should not be significant. Moreover, for stationary reasons, we expect λ_1 to be a number between zero and one. The degree of demand shifting across airports will be captured by the Spatial Autoregressive Coefficient λ_2, in Equation 2.

In addition to the spatial dependencies modeled here, there may also exist temporal dependencies in the sense that, for example, the price of the day before plays a role. Nearly all previous literature ignores both, spatial and temporal dependencies. This paper is the first to model the spatial dependencies. Additionally modeling the temporal dependencies goes beyond the scope of this paper.[3]

3.2 Market Definition and Spatial Weights

To estimate the model specified in Equations 1 and 2, we need to specify the spatial matrices W and M to construct the weighted average prices of the competing flights. Our measure of distance is the time difference between different departing flights. While some methods

[3] A recent paper that takes into account temporal dependencies is Escobari (2012).

of specifying the weight matrices include placing equal weights on all competitors within a critical distance, we will adopt a more realistic strategy and also consider a critical distance, but we place different weights based on the relative distance from competing flights. This is intuitively correct because the competition is expected to be greater for flights that have departure times that are closer. We use three different values of critical distance for the local market definition: 3 hours, 6 hours, and 12 hours. This will allow us to construct three different weighting matrices in the following way. Outside the critical distance the elements in the matrix are zero, while inside the critical distance each of the elements is defined as $w_{ij} = \omega_{ij}/\sum_{j=1}^{N}\omega_{ij}$. The value of ω_{ij} is calculated as $1/(1+d_{ij})$, where d_{ij} is the distance in departure times between flight i and flight j.

4 Results

The first set of estimates of Equation 1 using linear controls are presented in Table 2.[4] The numbers in parentheses are heteroskedasticity robust standard errors. The first three columns present the fixed effects estimates, while columns three through six present instrumental variable estimates with fixed effects. The coefficient of interest is the estimate of λ_1, from the variable W·Price. Different columns present estimates for different specifications of the weighting matrix W. Columns one and four present estimates of the price reaction function that take into account competing flight that are located within three hours. Columns two and five consider flights within 6 hours, while columns three and six flights within 12 hours.

The estimated coefficient of 0.309 means that the price of a given flight will increase by 30.9 cents if all the flights within 3 hours increase their prices by one dollar. Consistent with spatial competition models and the existence of demand shifting, the coefficients for 6 and 12 hours are subsequently larger. For example, the estimate in the third column implies that a given flight will increase its price by 55.7 cents if all the flights within 12 hours increase their prices by one dollar. Because the effects are larger when the market is more

[4] Even though we focus on one single route there is a substantial price dispersion — the average fare is 176 dollars and standard deviation is 139.1. Moreover, the highest fare is fourteen times the least expensive fare.

broadly defined, the difference between the estimated coefficients in the first three columns has a demand shifting connotation. If every flight within 3 hours increases its price, some demand will be diverted to flights that are more than three hours away; hence, the response is smaller (i.e., 30.9 cents). However, if every flight within twelve hours increases its price, there will not be diverted demand from the flights within 3 hours to flights located 3 to 12 hours away. Then the response is larger (i.e., 55.7 cents).

A second interpretation for these coefficients follows the diversion ratio explained in Pinkse et al. (2002) and used in Lee (2008). This interpretation comes from the assumption that n firms compete in a Bertrand-Nash simultaneous game, marginal cost is constant and firms face a downward sloping linear demand.[5] Then, the price reaction functions for each flight i can be written as:

$$\textsc{Price}_i = \mathrm{X}_i\beta + \lambda \sum_{j\neq i}^{n} w_{ij}\textsc{Price}_j + \mu_i \tag{3}$$

where $w_{ij} = d_{ij}/\sum_{j\neq i}^{n} d_{ij}$ and $\lambda = 1/2\sum_{j\neq i}^{n} d_{ij}$. The relative diversion ratios are given by w_{ij}, and $\sum_{j\neq i}^{n} w_{ij}\textsc{Price}_j$ is the weighted average price of i's competing flights. Therefore, when flight i increases its price, the coefficient of the competitors' weighted average price times two captures the fraction of sales lost.

Under this setting and based on the first column of Table 2, if a carrier increases its price, the flights within three hours will capture 61.8% (30.9% $\times$ 2) of the diverted demand. 81.4% percent of this demand will be captured by flights within 6 hours and the entire demand will be captured by flights within 12 hours. This diversion ratio does not consider the existence of outside goods, hence the possibility that the spatial autoregressive coefficient can be greater than 0.5.

A typical concern in the estimation of spatial models is the endogeneity of the spatial autoregressive term, W$\cdot$Price. In this paper we follow the approach described in Kelejian and Prucha (1998) and refined in Lee (2003). In our case this procedure involves a two-step method which yields an asymptotically optimal IV estimator when errors are i.i.d. In the first step we estimate Equation 1 via 2SLS using $\mathrm{H} = [\mathrm{X}, \mathrm{WX}, \mathrm{W}^2\mathrm{X}]$ as instruments. The estimator is $\hat{\theta}^{2SLS} = (\tilde{\mathrm{X}}'\mathrm{S}\tilde{\mathrm{X}})^{-1}\tilde{\mathrm{X}}'\mathrm{SP}$, where P is the vector of prices, $\tilde{\mathrm{X}} = [\mathrm{WP}, \mathrm{X}]$ is the matrix of explanatory variables, and $\mathrm{S} = \mathrm{H}(\mathrm{H}'\mathrm{H})^{-1}\mathrm{H}'$ is the weighting matrix. Using $\hat{\lambda}^{2SLS}$

[5] For a detailed derivation of the model, see Lee (2008).

and $\hat{\beta}^{2SLS}$ that are part of the vector $\hat{\theta}^{2SLS}$, the second step estimates an IV regression using instruments $\hat{\mathrm{Z}} = [\mathrm{X}, \mathrm{W}(\mathrm{I} - \hat{\lambda}^{2SLS})^{-1}\mathrm{X}\hat{\beta}^{2SLS}]$. That is, $\hat{\theta}^{IV} = (\hat{\mathrm{Z}}'\tilde{\mathrm{X}})^{-1}\hat{\mathrm{Z}}'\mathrm{P}$. These IV results are presented in columns 4 through 6 in Table 2. All spatial autoregressive terms are significant and larger than the ones when not controlling for endogeneity. The effect of competition appears to be larger and demand shifting is completely captured by flights within 3 hours. The regression estimates that consider nonlinearities in the control variables are presented in Table 3. The estimates of λ_1 are slightly smaller, but all highly statistically significant. The IV results still show that all demand is captured by flights within 3 hours based on the diversion ratio.

To measure demand shifting across airports, Table 4 shows the estimates of Equation 2. The coefficient on M·PRICE measures demand shifting across flights that depart from competing airports. The estimates are small and not significant, hence there is no evidence of demand shifting between the three airports that serve the New York City area.

5 Conclusions

This paper uses a unique panel data set of contemporaneous prices and inventories of competing flights for one particular route, New York City to Toronto. It is set to identify two particular effects. By estimating price reaction functions we capture the demand shifting across flights that compete using different departure times. We also tested for demand shifting for our particular route across the three airports that serve the New York City area.

After controlling for multiple sources of price dispersion that arise at the individual flight level, we are able to find that there is important demand shifting across flights. We find that the entire demand shifting is captured by competing flights that depart within 3 hours. We did not find any evidence of demand shifting between airports.

References

Alderighi, M., Cento, A., Nijkampm, P., Rietveld, P.: Second-degree price discrimination and inter-group effects in airline routes between European cities. Tinbergen Institute

Discussion Paper 2011-118/3 (2011)

Bilotkach, V.: Understanding price dispersion in the airline industry: Capacity constraints and consumer heterogeneity. In *Advances in Airline Economics*, Vol. 1, edited by Darin Lee. New York: Elsevier Science (2006)

Borenstein, S., Rose, N.L.: Competition and price dispersion in the U.S. airline industry. J. Polit. Economy **102**, 653-683 (1994)

Dana, Jr., J.D.: Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure. RAND J. Econ. **30**, 632-660 (1999)

Escobari, D.: Systematic peak-load pricing, congestion premia and demand diverting: Empirical evidence. Econ. Letters **103**, 59-61 (2009)

Escobari, D.: Dynamic pricing, advance sales, and aggregate demand learning in airlines. J. Ind. Econ. forthcoming (2012)

Gerardi, K.S., Shapiro, A.H.: Does competition reduce price dispersion? New evidence from the airline industry. J. Polit. Economy **117**, 1-37 (2009)

Kelejian, H.H., Prucha, I.R.: A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Finance Econ. **17**, 99-121 (1998)

Lee, L.F.: Best spatial two-stage least squares estimators for a spatial autoregressive disturbances. Econometric Rev. **22**, 307-355 (2003)

Lee, S.Y.: Three essays on spatial econometrics and empirical industrial organization, PhD Thesis, The Ohio State University (2008)

Pinkse, J., Slade, M.E., Brett, C.: Spatial price competition: A semiparametric approach. Econometrica **70**, 111-1153 (2002)

Prescott, E.: Efficiency of the natural rate, J. Polit. Economy **83**, 1229-1236 (1975)

Stavins, J.: Price discrimination in the airline market: The effect of market concentration. Rev. Econ. Statist. **83**, 200-202 (2001)

Table 1: Flights by Carrier and Date

	Friday (Dec 19)	Saturday (Dec 20)	Sunday (Dec 21)	Monday (Dec 22)	Tuesday (Dec 23)	Wednesday (Dec 24)	Total
American	12	8	11	12	13	11	67
Air Canada	19	9	13	20	20	16	97
Continental	8	5	7	8	8	6	42
Delta	4	4	4	4	4	3	23
Lan Chile	1	1	0	1	0	1	4
United	17	10	10	18	16	13	84
Total	61	37	45	63	61	50	317

Table 2: Across flights and linear controls

	Fixed Effects			Instrumental Variables		
Variables	3 hours	6 hours	12 hours	3 hours	6 hours	12 hours
W·PRICE	0.309*	0.407*	0.557*	0.546*	0.580*	0.934*
	(0.023)	(0.026)	(0.031)	(0.080)	(0.076)	(0.059)
SEATS	92.916*	89.553*	88.352*	157.336*	170.642*	176.890*
	(22.774)	(22.904)	(22.654)	(17.639)	(16.376)	(12.581)
ADVANCE	−2.847*	−2.565*	−2.346*	−2.304	−2.808**	−3.650*
	(0.317)	(0.301)	(0.318)	(1.486)	(1.418)	(1.037)
Within R-squared	0.199	0.211	0.226	0.186	0.204	0.211
Observations	4398	4398	4398	4398	4398	4398

Notes: The dependent variable is PRICE. The IV approach follows Lee (2003). Numbers in parentheses are heteroskedasticity robust standard errors. * significant at 1%; ** significant at 5%; *** significant at 10%. All specifications include flight-specific effects.

Table 3: Across flights and nonlinear controls

	Fixed Effects			Instrumental Variables		
Variables	3 hours	6 hours	12 hours	3 hours	6 hours	12 hours
W·Price	0.250*	0.340*	0.459*	0.518*	0.550*	0.834*
	(0.022)	(0.026)	(0.031)	(0.076)	(0.069)	(0.112)
Seats	295.696***	288.037***	237.774	567.733*	563.195*	474.242*
	(168.969)	(170.436)	(169.839)	(155.971)	(150.587)	(158.081)
Seats^2	−1289.13*	−1287.74*	−1147.56*	−1431.52*	−1414.62*	−1174.66*
	(311.384)	(313.319)	(310.992)	(295.769)	(294.713)	(301.969)
Seats^3	1177.43*	1177.96*	1074.56*	1056.71*	1158.10*	972.18*
	(181.422)	(182.629)	(181.821)	(160.351)	(174.940)	(182.741)
Advance	−10.336*	−12.981*	−11.929*	−9.257**	−11.841*	−12.085*
	(3.601)	(3.606)	(3.572)	(4.060)	(4.014)	(4.055)
Advance^2	0.339**	0.435*	0.387**	0.340**	0.416**	0.359**
	(0.162)	(0.162)	(0.161)	(0.169)	(0.170)	(0.168)
Advance^3	−0.004***	−0.005**	−0.004**	−0.004***	−0.005**	−0.004***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Within R-squared	0.251	0.263	0.271	0.229	0.249	0.250
Observations	4398	4398	4398	4398	4398	4398

Notes: The dependent variable is Price. The IV approach follows Lee (2003). Numbers in parentheses are heteroskedasticity robust standard errors. * significant at 1%; ** significant at 5%; *** significant at 10%. All specifications include flight-specific effects.

Table 4: Across airports and nonlinear controls

Variables	3 hours	6 hours	12 hours
W·PRICE	0.245*	0.338*	0.464*
	(0.024)	(0.028)	(0.033)
M·PRICE	0.018	0.005	−0.018
	(0.025)	(0.033)	(0.029)
SEATS	300.562***	288.778***	234.743
	(171.750)	(170.875)	(170.104)
SEATS^2	−1292.45*	−1288.24*	−1143.76*
	(311.434)	(312.680)	(311.651)
SEATS^3	1177.05*	1177.98*	1072.60*
	(181.364)	(182.633)	(181.797)
ADVANCE	−10.597*	−13.043*	−11.756*
	(3.617)	(3.623)	(3.595)
ADVANCE^2	0.348**	0.437*	0.381**
	(0.163)	(0.163)	(0.162)
ADVANCE^3	−0.004***	−0.005**	−0.004**
	(0.002)	(0.002)	(0.002)
Within R-squared	0.251	0.263	0.271
Observations	4398	4398	4398

Notes: The dependent variable is PRICE. Numbers in parentheses are heteroskedasticity robust standard errors. * significant at 1%; ** significant at 5%; *** significant at 10%. All specifications include flight-specific effects.