2014

The choice of airport, airline, and departure date and time: Estimating the demand for flights

Diego Escobari

Cristhian Mellado

The Choice of Airport, Airline, and Departure Date and Time: Estimating the Demand for Flights

Escobari, Diego and Mellado, Cristhian

The University of Texas - Pan American, Universidad Catolica de la Santisima Concepcion

July 2013

The Choice of Airport, Airline, and Departure Date and Time: Estimating the Demand for Flights*

Diego Escobari†　　Cristhian Mellado‡

July 2013

Abstract

This paper estimates the demand for flights in an international air travel market using a unique dataset with detailed information not only on flight choices but also on contemporaneous prices and characteristics of all the alternative non-booked flights. The estimation strategy employs a simple discrete choice random utility model that we use to analyze how choices and its response to prices depend on the departing airport, the identity of the carrier, and the departure date and time. The results show that a 10% increase in prices in a 100-seat aircraft throughout a 100-period selling season decreases quantity demanded by 7.7 seats. We also find that the quantity demanded is more responsive to prices for Delta and American, during morning and evening flights and that the response to prices changes significantly over different departure dates.

Keywords: Airline demand, discrete choice, flight choice, demand estimation.

JEL Classifications: R41, C25, L93.

---

*We thank Stephanie Reynolds and Sang-Yeob Lee for their work with the data.

†The University of Texas - Pan American, Department of Economics and Finance, 1201 West University Drive, Edinburg, TX 78541, escobarida@utpa.edu, http://faculty.utpa.edu/escobarida

‡Universidad Católica de la Santísima Concepción, Department of Economics, Alonso de Ribera 2850, Concepción, Chile, and The University of Texas - Pan American, Department of Economics and Finance, 1201 West University Drive, Edinburg, TX 78541, crmelladocid@utpa.edu

1 Introduction

The main goal of empirical studies of differentiated product markets is the estimation of the demand and the modeling of choices. The common practice is to estimate random utility models of consumer demand using discrete choice models. A typical problem in this type of estimation is the difficulty in obtaining the data on choices and on all the available options. This paper proposes using a unique dataset on prices and transactions to estimate discrete choice models that explain individual choices at the flight level. In our data we observe not only the flight choices and the corresponding prices but also the prices and characteristics of all the non-booked alternative flights. We are able to estimate how air travel demand changes depending on the departure date, the departure time, the identity of the carrier and the departing airport. The passenger choice model presented here can be vital to the development and the assessment of new pricing strategies, capacity choices, or route entry/exit decision. Moreover, it opens the door to a large number of other discrete choice models that can be estimated using similar datasets. These include, for example, models to forecast demand, estimation of market power, cross-price elasticities, single agent dynamic models, or dynamic games.

To the best of our knowledge this is the first study that estimates the itinerary choice (i.e., flight choice) in a revealed preference setting where information on choices and all the alternative flights is available. The information on all options is important because this is part of the information set available to customers when they make their choices. This allows an easy construction of the representative utility in a random utility model. When not all the available alternatives are observed the estimation is complicated due to the lack of information on the arrival rates. For example, some unobserved customers can arrive and purchase from a seller that is not in the data.[1]

The airline industry has already demonstrated to be a popular place for the estimation of discrete choice models using aggregate data from the U.S. Department of Transportation. However, these data is too aggregate to estimate the individual choices. Estimating discrete choice models of demand employs aggregate market level data, consumer level data or

[1]When there is missing data one potential solution is to use the techniques described in Newman et al.(2012).

both. Aggregate level data usually has aggregate quantity, prices, consumer characteristics, market size and sometimes the distribution of demographics. It has the advantage that in most cases it is easier to get. Consumer-level data has individual choices, prices and characteristics of all options. The distribution of demographics is optional. It has the advantage that it is more detailed than aggregate data, but it is more difficult to obtain. We use posted prices and inventory changes following a similar collection strategy as Escobari (2012) and Escobari and Lee (2012).

In this paper we focus on the international market between the three big airports serving New York City (Newark Liberty, John F. Kennedy and La Guardia) and the main airport in Toronto (Toronto Pearson International). The data covers all the 317 flights from the six carriers that served this city pair between December 19 and 24, 2008. We cover the advance sales during forty days leading to the departure dates. Counting the choices and the available alternatives we have more than half a million observations in our dataset. The results from the estimation of our random utility framework show that a 10% increase in prices throughout a 100-period selling season decreases quantity demanded by 7.7 seats in a 100-seat flight. We also find that demand is greater closer to departure. When allowing the response of quantity demanded to prices to change with the identity of the carrier we found that Delta Airlines had the most price responsive demand followed by American Airlines and Air Canada. Additional results show that less responsive demand is associated with departure dates that have more congestion and higher prices and that the responsiveness to prices varies significantly across departure dates. Finally, demand is more responsive to prices for flights that depart from John F. Kennedy and during morning flights.

There is important related literature in airlines that uses discrete choice estimations aimed at modeling consumer choices. Pels et al.(2001) and Hess and Polak (2005) use data from an airline passenger survey to estimate various logit models of airport and airline choice for the San Francisco Bay area. Ashiabor et al.(2007) use the 1995 American Travel Survey to forecast travel demand, while Proussaloglou and Koppelman (1999) and Wen and Lai (2012) also use survey data to estimate the choice of carrier. It is important to notice that survey data basically creates trip scenarios to simulate the booking process. These surveys record stated preferences and not revealed preferences based on actual choices that result in transactions. Our data comes from the actual behavior of the interaction between sellers and buyers recorded by sales and posted prices. Carrier (2008) estimates itinerary choice using booking data, but does not have the non-booked travel alternatives as we do. Using revealed preferences from a single major European airline and stated preferences, Atasoy and Bierlaire (2012) present an itinerary choice model. Related literature on airlines that used the most common Airline Origin and Destination Survey (DB1B) from the U.S. Department of Transportation includes Berry (1992) who estimates a structural discrete-choice model of entry for the airline industry. Berry and Jia (2010) present a structural model and estimate the impact of demand and supply changes on profitability during the turmoil in the industry in the early 2000s, while Ciliberto and Tamer (2009) used a partially identified entry model to investigate the heterogeneity in carriers profits.[2]

While we propose using posted prices, inventory changes and discrete choice models to explain demand side behavior, posted prices data has been very popular for the estimation of pricing strategies and supply side behavior. Stavins (2001) uses posted prices from the Official Airline Guide to find that price dispersion attributed to ticket restrictions increases with competition. More recently McAfee and te Velde (2007) looks at price dynamics, van Eggermond et al.(2007) at travelers itinerary in European markets, Mantin and Koo (2009) study dynamic price dispersion, and Alderighi (2010) explains fare dispersion. Bilotkach et al.(2011) show empirically how yield management is effective in raising a flight's load factor, while Bilotkach and Rupp (2012) study the intertemporal profile and the role of low-cost carriers and differences across online distributors. The intertemporal profile of fares in also studied in Bergantino and Capozza (2012), who find a J-curve and in Alderighi et al.(2012) who find a U-shape and that fares increase with occupancy rate. Escobari (2012) estimates a dynamic demand equation and a dynamic supply equation that jointly explain the dynamics of fares and sales as the departure date nears, Escobari and Lee (2012) estimate price reaction functions to capture the interaction between flights, and Escobari et al.(2013) shows empirically how airlines dynamically price discriminate.[3]

The organization of the rest of the paper is as follows. Section 2 explains the collection of the data and presents the summary statistics. The explanation of the discrete choice

[2] See also Armantier and Richard (2008), and more recently Gedge et al.(2013).

[3] On the theoretical side, Deneckere and Peck (2012) present a theory to explain price posting in a multiple period version of the models presented in Dana (1998) and Dana (1999).

empirical model is presented in Section 3. Section 4 presents the results, while Section 5 discusses possible extensions. Section 6 concludes.

2 Data

The data for this paper was collected from the online travel agency Expedia.com following a similar strategy as in Escobari (2012) and Escobari and Lee (2012). We not only have information on all contemporaneous posted prices at different points in time for all available options (i.e., flights) but also on seat inventory levels. We use the changes in inventories to identify a sale. Hence our data replicates the information displayed to individuals who buy tickets online and records their choices. To control for various sources of price dispersion and demand variation across customers we focus on the lowest available one-way non-stop economy class nonrefundable posted fare. Looking only at one-way non-stop tickets helps define a single inventory at each price and helps control for tickets sold as part of round-trips or longer itineraries.

As in Escobari (2012), even if one-way tickets are a small fraction of the overall tickets sold, our observed prices are relevant as long as the carriers adjust these prices based on the current inventory levels. Moreover, sales (obtained as inventory changes) can be the result of tickets sold at prices different than the one-way prices - for example, as a round-trip ticket where one of the legs is in our sample. Then our demand estimation is also capturing the demand of round-trip tickets if one-way tickets are always priced half of the round-trip tickets, which is the standard assumption in the airline pricing empirical literature.[4] In addition, to make the problem tractable we focus on a single city pair, New York City to Toronto which already generates over half-a-million observations. Because there are three big airports that serve New York City, we collected the data for all three airports, Newark Liberty International Airport (EWR), John F. Kennedy International Airport (JFK), and La Guardia Airport (LGA). The only airport in Toronto that we consider is the Toronto Pearson International Airport (YYZ), which is the only big airport that serves this city.

We have sales and pricing information on all the flights that departed between December 19 and December 24, 2008. Moreover, we keep record of the prices and inventory changes

[4] See for example Borenstein and Rose (1994, p. 677), and Gerardi and Shapiro (2009, p. 5).

every three days between 40 days to departure and 1 day to departure for all these flights. Overall this included 317 flights from American Airlines, Air Canada, Continental, Delta, Lan Chile, and United, with 10,708 tickets sold during our period of study. The details of the 317 flights by carrier and by departure date are presented in Table 1, which is reproduced from Escobari and Lee (2012).[5] Each time a ticket was sold, we recorded the corresponding price and the prices of all competing flights for the same day of departure. This makes our data replicate the information available to the buyer as well as the same structure as required for the estimation of discrete choice models. Each consumer at the time of arriving to buy a ticket observes all posted prices and chooses the flight that gives him the highest utility. On average we have that for each recorded ticket transaction we there are 52 competing flights. For example, if a traveler buys a ticket from United to fly on December 24, we also record the contemporaneous posted prices for the other 49 competing flights as detailed in Table 1.

[Table 1, about here]

In addition to sales and prices that allows estimating airline demand, the data is interesting because allows demand comparisons between departing airports, departure dates, carriers, and departure times. Table 2 reports the summary statistics for these variables. The first four columns report the typical statistics, while columns 5 through 8 reports the mean, standard deviation, minimum and maximum prices for each of the classifications dictated by the dummy categories. The figures show that the dominant carrier in this city pair is Air Canada with 30.9% of the flights, followed by United with 25.8% and by American Airlines with 20.8%. Moreover, there is substantial price dispersion in the sample. The lowest priced ticket at US$ 65 is more than 16 times cheaper than the most expensive ticket. The busiest airport is La Guardia, while most flights depart in the morning. It is interesting to notice that while the overall average price is US$ 169.80 (column 1), the average price at which a transaction occurred is US$ 156.93 (column 5).

[5]While the data collection process in Escobari and Lee (2012) is the same as in this paper, here the structure and approach is very different. Escobari and Lee (2012) have 4,398 observations on posted prices. Here we have 560,244 observations that keep track of the posted prices of all the available flights for every time we observe a sale.

[Table 2, about here]

There are important advantages in the structure of the collection of the data and in focusing on this particular city pair. First, direct flights between New York City and Toronto take only one hour and a half, hence it is reasonable to think that combination of flights connecting this city pair with one or more stops are not a desired alternative for travelers. Moreover, focusing on non-stop flights and one-way fares is useful to control for fare differences associated to round-trip tickets and open jaws. For example, these tickets are usually associated with Saturday-night-stay restrictions or minimum-, and maximum-stay restrictions. This would involve tickets of a significantly different quality. Selecting the least expensive available non-refundable ticket is important to control for the existence of more expensive refundable tickets that are also available for purchase at different points prior to departure. Finally, focusing on economy class tickets controls for some consumer's heterogeneity as some higher valuation consumers may want to buy first-class tickets. We consider refundable and first class tickets to be of a significantly different quality.

3 Empirical Model

Consider the following random utility model framework in which travelers are assumed to be utility maximizers. We adapt the model of Chapter 2 in Train (2002) for our setting.[6] Let traveler n face a choice of traveling in any of J different flights. Notice that these J flights are all the available flight options shown by the online travel agency. The utility of individual n obtained from alternative $j \in J$ is U_{nj}. This level of utility is known by the traveler but not by the econometrician. We assume that the traveler already decided to fly over other transportation alternatives (e.g., driving or renting a car). Hence, he will choose the flight that gives him the highest level of utility. That is, he will choose flight i if and only if $U_{ni} > U_{nj}$ for all $j \neq i$. Even thought the econometrician cannot observe the utility levels, some of the flights' attributes including the price can be observed. We label them as x_{nj} and p_n respectively. We will relate these observed factors to the traveler's utility with

[6]Talluri and van Ryzin (2004) study a similar buyers' choice behavior from a revenue management perspective. For a detailed discussion of discrete choice models of airline demand see Garrow (2010).

the function $V_{nj} = V(x_{nj}, p_{nj})$, which is called the representative utility. Some attributes in x_{nj} can include, for example, departure time, departure date, departing airport, and identity of the carrier.

From the econometrician's view point, utility levels U_{nj} contain some random unobserved component that make $U_{nj} \neq V_{nj}$. We then write $U_{nj} = V_{nj} + \varepsilon_{nj}$, with ε_{nj} being the stochastic component of utility. We write $f(\varepsilon_n)$ as the joint density of the random vector $\varepsilon_n = \{\varepsilon_{n1}, \varepsilon_{n2}, \ldots, \varepsilon_{nJ}\}$. Then the probability that traveler n chooses alternative i is given by,

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \quad \forall j \neq i), \\ &= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall j \neq i), \\ &= \text{Prob}(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \quad \forall j \neq i). \end{aligned} \tag{1}$$

Assume that ε_{nj} is distributed independent and identically distributed extreme value. Hence, the distribution of each unobserved component of utility is:

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}, \tag{2}$$

with the cumulative distribution being $F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$. Because the difference between two extreme value variables is logistic we have

$$F(\varepsilon_{nj} - \varepsilon_{ni}) = \frac{e^{\varepsilon_{nj} - \varepsilon_{ni}}}{1 + e^{\varepsilon_{nj} - \varepsilon_{ni}}}. \tag{3}$$

From Equation (1) if ε_{ni} is taken as given, the cumulative probability distribution for each ε_{nj} evaluated at $\varepsilon_{ni} + V_{ni} - V_{nj}$ based on Equation (3) is $e^{-e^{-\varepsilon_{ni} + V_{ni} - V_{nj}}}$. Following the assumption of independence, this cumulative distribution over all $j \neq i$ is the product of the individual cumulative distributions:

$$P_{ni} | \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-\varepsilon_{ni} + V_{ni} - V_{nj}}}. \tag{4}$$

Because ε_{ni} is not given, we need to take the integral of $P_{ni} | \varepsilon_{ni}$ over all values of ε_{ni} weighted by the density of ε_{ni}:

$$P_{ni} = \int \Big(\prod_{j \neq i} e^{-e^{-\varepsilon_{ni} + V_{ni} - V_{nj}}} \Big) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \tag{5}$$

that has the integral equal to

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{6}$$

which is the equation for the logit choice probability. Because we have a panel the utility that the traveler n obtains from buying a ticket on flight j at time t is given by,

$$U_{njt} = \beta a_{njt} + \alpha_j p_{njt} + \mu_j + \varepsilon_{njt}. \tag{7}$$

In Equation (7) we model the systematic component of utility as a linear function of the parameters, $V_{njt} = \beta a_{njt} + \alpha_j p_{njt} + \varepsilon_j$. β and α_j are coefficients to be estimated and the variables a_{njt} and p_{njt} are the number of days in advance the ticket was bought and the price. Our main variable of interest is α_j which captures the marginal utility of a price increase. Of course we expect α_j to be negative, hence $-\alpha_j$ is the disutility of a price increase. The key element in which flights are differentiated are all time invariant and captured by μ_j, the time-invariant fixed effect specific to flight j. This one controls for observed and unobserved flight specific characteristics such as departure time, departure date, identity of the carrier, distance between the cities, or the carrier's managerial capacity.

As a first approach we fix $\alpha_j = \alpha \ \forall j$ but then we allow α_j to change with time invariant flight specific characteristics. Then the systematic component of utility will be modeled as $V_{njt} = \beta a_{njt} + (\delta' x_{nj}) \cdot p_{njt} + \mu_j$ where we just allow α_j in Equation (7) to be $\alpha_j = \delta' x_{nj}$. The vector of variables x_{nj} includes departure time, departure date, the identity of the carrier and the departing airport. This is alternative specification in helpful to determine how demand for a flights changes with these characteristics. There are some assumptions for the validity of the estimation of Equation (7) that discuss in detail along with the results in the next section.

4 Results

In this section we present the discrete choice logit estimates of the demand. For comparison purposes we also report the linear pooled OLS as well as the linear fixed effects estimates. Table 3 present the base model of Equation (7) where a_{njt} is relabeled as ADVANCE, the number of days prior to departure fares and the sale was recorded. For p_{njt} we use

LogPrice, the natural logarithm of Price. For simplicity we omit the subscripts njt from the names of the variables. The negative and statistically significant coefficients on Advance in columns 1 and 3 suggest that sales (demand) increases closer to departure. However, this result is not robust across specifications as the fixed effects estimate finds a non-significant coefficient. Notice that the reported logit coefficients are the marginal effects obtained when evaluating the other regressors at their mean levels.

It is important to notice that LogPrice in the estimation is potentially endogenous. Endogeneity arises if there is correlation between p_{njt} and the unobserved $\mu_j + \varepsilon_{njt}$. The most common cause of this correlation is if the carrier sets prices knowing more about the error term than the econometrician. Escobari (2012) controls for potential endogeneity in a dynamic setting using internal instruments. Here we control for the potential endogeneity that arises due to correlation between p_{njt} and ε_j using flight fixed effects. For ε_{njt} we assume that it is uncorrelated with p_{njt}. This assumption is reasonable given that in Escobari (2012) the point estimates in the within specification that consider fare as exogenous is virtually the same as the point estimate that treats prices as endogenous.[7] An alternative approach would be to follow the methods developed in Berry et al.(1995). However, this approach still needs exogenous instruments which in most cases use supply side variables.

[Table 3, about here]

The negative point estimate for the coefficient on LogPrice is statistically significant at a 1% and it is consistent with a downward sloping demand. The coefficient in column 3 indicates that when the price of a ticket increases by 10%, quantity demanded decreases by 0.070 seats in a 100-seat aircraft.[8] This is a reasonable value for the estimate because this is what quantity demanded decreases each period prior to departure. If prices are 10% higher throughout the selling season and there are 100 periods, then this particular flight will sell 7 seats less. The point estimate in column 4 that controls for flight fixed effects suggest a

[7]See Table IV, columns (2) and (8) in Escobari (2012).

[8]This calculation follows that (-0.007/100)(%ΔPrice), which is the marginal effect given that LogPrice on the right-hand side. A 10% increase in price (%ΔPrice=10) decreases the dependent variable by 0.0007 that in a 100-seat aircraft is 0.07 seats.

slightly larger. This coefficient also captures the disutility of a price increase. Notice that while we have a number of variables reported in Table 2 besides Advance that can be part of Equation (7) (e.g., carrier identifier, airport identifier, departure time and departure date), we cannot separately identify the marginal effects of those variables because they are perfectly collinear with the flight fixed effects α_j. However we can estimate interaction terms between those variables and LogPrice to see how the slope of the demand changes.

[Table 4, about here]

Table 4 estimates a model in which the slope α is allowed to change with the identity of the carrier. Focusing on the last column of the table we can observe that the carrier with the most responsive demand is Delta, followed by American Airlines and Air Canada. The least responsive demand is the one for United. Notice that these differences have a relatively big economic importance. For example, all else constant, a 10% price increase throughout a 100-period selling season in a 100-seat aircraft decreases the quantity demanded by 18 seats in Delta flight (the largest response) and by slightly less than one seat in a United flight (the smallest response).

More inelastic demands are usually associated with higher market power and the ability to charge higher prices. Interestingly, Delta that is the carrier with the less responsive demand is also the one with the lowest presence in this city pair. As column 1 in Table 2 shows, only 7.32% of the flights in this city pair belong to Delta. However, Delta also charges the highest average prices ($234.36) as reported in column 5 of Table 2. Air Canada, who is the dominant carrier in this city pair with 30.9% of the flights is not the carrier that charges the highest average fares and does not have the most responsive (or nonresponsive) demand. It is difficult to infer about any causality from these results because from the view point of the econometrician, fares and market presence are jointly determined.

[Table 5, about here]

Table 5 shows how the effect of price on quantity demanded changes with the departure date. Demand is less responsive the Friday (December 19) before Christmas and is it more responsive two days before Christmas. With the exception on December 19, there is not a big economically significant difference in the response across the rest of the departure dates - all coefficients lie between 0.011 and 0.015. On December 21, a 10% increase in prices throughout a 100-period selling season in a 100-seat aircraft decreases quantity demanded by 11.1 seats while on December 23 the effect is only 15.4 seats. Compared with the other departure dates on Tuesday December 23 the demand is relatively more responsive. This is interesting because that Tuesday is the day where there is a large number of scheduled flights, 61 (see Table 1), which corresponds to 21.4% of the flights in the sample (see column 1 in Table 2). An interesting point in Table 2 is the link between higher fares and congestion. For example, the day in which the least number of flights were scheduled is Saturday December 20 with only 37 flights. This is also the departure date with the highest average fares ($279.85, see column 5, Table 2). Higher fares associated with more congestion known ex-ante is evidence of systematic peak-load pricing, as previously documented in Escobari (2009). It would be reasonable to observe that a more responsive demand is associated with more congestion, but the estimates in Table 5 show that there is little evidence that this is the case.

[Table 6, about here]

The regression estimates in Table 6 are presented to address the role of the departing airport on the relationship between prices and quantity demanded. Column 4 shows that the point estimates for Newark of for La Guardia are nearly the same, while the response in flights that depart from John F. Kennedy is much smaller. The null that the coefficients between Newark and La Guardia are the same is rejected at the any reasonable significance levels. Notice that the last three rows in the table report the p-values of all the null hypotheses that test for pair-wise differences in the coefficients across airports. The last two rows show that there response in the John F. Kennedy airport is significantly different that in the other two airports. A 10% increase in prices throughout a 100-period selling season in a 100-seat aircraft decreases quantity demanded by 17.8 seats in a flight departing from the John F. Kennedy airport, while for the La Guardia and for Newark this figure is 57.9 seats and 66.5 seats respectively.

[Table 7, about here]

The final set of estimates is presented in Table 7. Here the goal is to assess the role of the departure time on the effect that prices have on quantity demanded. We divide flight departure times in three, morning if the flight departs before noon, afternoon if the flight departs between noon and 5:00 p.m., and evening if the flight departs after 5:00 p.m. The logit estimates in the last column that control for flight specific characteristics show that quantity demanded is less responsive for flights departing in the afternoon and about equally responsive for flights departing either in the morning or in the evening. The last three rows reports the p-values for various null hypotheses that the coefficients are pair-wise equal.

The logit estimator used in this paper has some limitations. As explained in Train (2002), this logit estimator can represent systematic taste variation but not random taste variation. Logit can approximate average tastes fairly well even when tastes are random, but a probit or a mixed logit model may be better at including random taste variation. A second potential limitation is that the logit model has the independence of from irrelevant alternatives (IIA) property, which means that the ratio of the probabilities of two alternatives does not depend on any other alternative. Finally, the logit cannot handle unobserved factors that are correlated over time. The main goal in this paper is to illustrate the use of our unique data to estimate flight choice models, but other logit estimators can also be used as extensions to this research that overcome some of the limitations outlined above. In the next section we provide some examples for further research using similar datasets.

5 Potential Extensions

In this paper we estimated a simple random utility model, but our data and the discrete choice modeling approach used in this paper can be extended to a number of settings. The most obvious is to use these model estimates to forecast flight-level demand. This is or particular importance for carriers because they use forecasted demand to schedule flights.

Additional potential extensions include measuring market power at the route level and merger evaluation using the methods proposed in Nevo (2001) or welfare from new flights following techniques for the introduction of the minivan in Petrin (2002). Discrete choice models can also be used to estimate pricing strategies in monopoly routes using single agent dynamics as in Rust (1987), Hotz and Miller (1993), or Aguirregabiria and Mira (2002). In these models agents are forward looking and maximize intertemporal payoffs. When more than one carrier serves a route, the approach can as well follow Bajari et al.(2006) or Aguirregabiria and Mira (2007) to estimate dynamic discrete games to model and estimate the interaction between flights. The well-known challenges in these dynamic estimations are the large number of agents, choices and states, and the existence of multiple equilibria, which means an important computational burden. Igami (2013) overcomes these challenges by modeling a small number of state spaces and choice sets to estimates a dynamic model via maximum likelihood using the nested fixed-point algorithm of Rust (1987). Some of these topics have already been addresses using posted prices but without discrete choice models. Escobari (2012) estimates dynamic demand and dynamic pricing equations in a setting where agents are forward looking. Moreover, Escobari and Lee (2012) estimate price reaction functions to model the interaction between agents.

6 Conclusion

The choice of transportation mode is the most widely used example to illustrate random utility models. This follows from the seminal work of Daniel McFadden (1974) that estimates the trip mode choice in a study of travel demand. On this line modeling the choice of flights has also been a popular research topic; however, obtaining the appropriated data proved to be a difficult task. Previous studies either worked with stated preferences based on survey data or with revealed preference based on data from a single seller. In this paper we work with revealed preferences based on data on all flights from the international air travel market between New York City and Toronto. Our study combines two key pieces of information. First, we have posted prices at each time period and for all the available booked and non-booked flights. This records the menu of options available to the buyers. Second, we observe changes in inventory levels which allow us to identify choices. The panel structure of the data with multiple transactions per flight at various points prior to departure allows us to control for unobserved flight specific characteristics.

Our demand estimates show that in a 100-seat flight a 10% increase in prices throughout a 100-period selling season reduces quantity demanded by 7.7 seats. We further inquire how this figure changes based on key sources of product differentiation in airline markets - the departing airport, the identity of the carrier and the departure date and time. We find that quantity demanded is more responsive to prices for Delta flights, followed by flights from American and Air Canada. Moreover, there are significant differences in this responsiveness across departure dates and departing airport. Demand is less responsive for departures at John F. Kennedy than for departures at La Guardia or Newark. Finally, quantity demanded is more responsive to prices for departures in the morning and in the evening when compared to departures in the afternoon. We discuss the validity of our assumptions and highlight potential areas for future research.

References

Aguirregabiria, V., and P. Mira, 2002. Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models. *Econometrica* 70, 1519-1543.

Aguirregabiria, V., and P. Mira, 2007. Sequential estimation of dynamic discrete games. *Econometrica* 75, 1-53.

Alderighi, M., 2010. Fare dispersions in airline markets: A quantitative assessment of theoretical explanations. *Journal of Air Transport Management* 16, 144-150.

Alderighi, M., M. Nicolini, and C.A. Piga, 2012. Combined effects of load factors and booking time on fares: Insights from the yield management of a low-cost airline. Nota di Lavoro, Fondazione Eni Enrico Mattei.

Armantier, O., and O. Richard, 2008. Domestic Airline Alliances and Consumer Welfare. *Rand Journal of Economics* 39, 875-904.

Ashiabor S., H. Baik, and A. Trani, 2007. Logit models for forecasting nationwide intercity travel demand in the United States. *Transportation Research Record* 2007, 1-12.

Atasoy B., and M. Bierlaire, 2012. An air itinerary choice model based on a mixed RP/SP dataset. Technical Report TRANSP-OR 120426, Transport and Mobility Laboratory, Ecole Polytechnique Federale de Lausanne, 2012.

Bajari, P., L. Benkard, and J. Levin, 2006. Estimating dynamic models of imperfect competition. *Econometrica* 75, 1331-1370.

Bergantino, A.S., and C. Capozza, 2012. Airline pricing behaviour under limited intermodal competition. Working Paper SIET.

Berry, S., 1992. Estimation of a model of entry in the airline industry. *Econometrica* 60, 889-917.

Berry, S., and P. Jia, 2010. Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics* 2, 1-43.

Berry, S., J. Levinsohn, and A. Pakes, 1995. Automobile prices in market equilibrium. *Econometrica* 63, 841-890.

Bilotkach, V., A. Gaggero, and C.A. Piga, 2011. Airline pricing under different market conditions: Evidence from European low-cost carriers. Rimini Centre for Economic Analysis, Working Paper 11-47.

Bilotkach, V., and N.G. Rupp, 2012. A Guide to booking airline tickets online, in James Peoples (ed.) Pricing Behavior and Non-Price Characteristics in the Airline Industry (Advances in Airline Economics, Volume 3), Emerald Group Publishing Limited, pp.83-105.

Carrier, E. 2008. Modeling the choice of an airline itinerary and fare product using booking and seat availability data, PhD thesis, Massachusetts Institute of Technology.

Ciliberto, F. and E. Tamer, 2009. Market structure and multiple equilibria in airline markets. *Econometrica* 77, 1791-1828.

Dana Jr., J.D., 1998. Advance-purchase discounts and price discrimination in competitive markets. *Journal of Political Economy* 106, 395-422.

Dana Jr., J.D., 1999. Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure. *Rand Journal of Economics* 30, 632-660.

Deneckere, R. and J. Peck, 2012. Dynamic competition with random demand and costless search: A theory of price posting. *Econometrica* 80, 1185-1247.

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

Escobari, D., 2012. Dynamic pricing, advance sales, and aggregate demand learning in airlines. *Journal of Industrial Economics* 60, 697-724.

Escobari, D. and S-Y. Lee. 2012. Demand Shifting across Flights and Airports in a Spatial Competition Model. *Letters in Spatial and Resource Sciences* 5, 175-183.

Escobari, D., N. Rupp., and J. Meskey, 2013. Dynamic price discrimination in airlines. Mimeo, East Carolina University.

Gedge, G., J.W. Roberts, and A. Sweeting, 2013. A model of dynamic limit pricing with an application to the airline industry. Mimeo, Duke University.

Garrow, L.A., 2010. Discrete Choice Modeling and Air Travel Demand: Theory and Applications. Ashgate Publishing Limited.

Hess, S., and J.W. Polak, 2005. Mixed logit modelling of airport choice in multi-airport regions. *Journal of Air Transport Management* 11, 59-68

Hotz, J., and R.A., Miller, 1993. Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies* 60, 497-529.

Igami, M., 2013. Estimating the Innovator's Dilemma: Structural Analysis of Creative Destruction. Mimeo, Yale University.

Mantin, B. and B. Koo, 2009. Dynamic price dispersion in airline markets. *Transportation Research Part E* 45, 1020-1029.

McAfee, R. P. and V. te Velde, 2007. Dynamic pricing in the airline industry. In Hendershott T. J. (ed.), Handbook of Economics and Information Systems, Vol. 1, (Elsevier Science, New York, New York, U.S.A.).

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior, In Frontiers in Econometrics, ed. by P. Zarembka, New York: Academic Press, 105-142.

Nevo, A., 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69, 307-342.

Newman, J.P., L. Garrow, M. Ferguson, T.L. Jacobs, and H. Purnomo, 2012. Estimation of choice-based models using sales data from a single firm. Mimeo, Georgia Institute of Technology.

Petrin, A., 2002. Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy* 110, 705-729.

Proussaloglou, K, and F.S. Koppelman, 1999. The choice of air carrier, fight, and fare class. *Journal of Air Transport Management* 5, 193-201.

Rust, J., 1987. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55, 999-1033.

Stavins, J., 2001. Price discrimination in the airline market: The effect of market concentration. *The Review of Economics and Statistics* 83, 200-202.

Talluri, K. and G. van Ryzin, 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50, 15-33.

Train, K., 2002. Discrete Choice Methods with Simulation, Cambridge University Press, New York.

Van Eggermond M.A.B., N. Schuessler, and K.W. Auxhausen, 2007. Consumer choice behavior and strategies of air transportation service providers. In Proceedings 7th Swiss Transport Research Conference (Monte Verita, Ascona, Switzerland).

Wen, C-H. and S-C. Lai, 2010. Latent class models of international air carrier choice. *Transportation Research Part E* 46, 211-221.

Table 1: Flights by carrier and date

	Fri Dec 19	Sat Dec 20	Sun Dec 21	Mon Dec 22	Tue Dec 23	Wed 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: Summary statistics

	Main Variable				Price			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Mean	SD	Min	Max	Mean	SD	Min	Max
Price	169.8	133.3	65	1,075				
Advance	19.00	12.43	1	40				
Bought	0.0191	0.137	0	1	156.93	111.87	65	1075
Carriers:								
American	0.208	0.406	0	1	192.38	194.72	76	1075
Air Canada	0.309	0.462	0	1	144.57	69.15	81	736
Continental	0.138	0.345	0	1	155.21	114.51	77	1001
Delta	0.0732	0.261	0	1	234.36	203.77	105	953
Lan Chile	0.0129	0.113	0	1	130.45	25.84	123	220
United	0.258	0.438	0	1	173.42	105.25	65	1008
Departure Dates:								
December 19	0.247	0.431	0	1	142.21	95.33	76	944
December 20	0.061	0.239	0	1	279.85	241.12	65	1075
December 21	0.108	0.311	0	1	180.72	98.35	81	738
December 22	0.231	0.422	0	1	128.38	38.72	76	472
December 23	0.214	0.41	0	1	193.73	159.52	81	1075
December 24	0.138	0.345	0	1	194.51	160.96	81	1075
Airports:								
Newark Liberty	0.356	0.479	0	1	150.73	101.27	65	1001
La Guardia	0.515	0.500	0	1	171.87	133.96	76	1075
John F. Kennedy	0.129	0.336	0	1	214.46	186.27	87	1075
Departure Times:								
Morning	0.447	0.497	0	1	142.06	95.94	65	1075
Afternoon	0.378	0.485	0	1	192.68	162.87	77	1075
Evening	0.175	0.380	0	1	191.55	130.96	81	953

Notes: The sample size is 560,244.

Table 3: Demand Estimates, base model

	(1)	(2)	(3)	(4)
VARIABLES	Pooled	FE	Logit	Logit
LogPrice	-0.00642***	-0.0113***	-0.00704***	-0.00771***
	(0.000416)	(0.000465)	(0.000459)	(0.000501)
Advance	-0.000154***	0.000832	-0.000151***	-0.000166***
	(1.53e-05)	(0.00106)	(1.50e-05)	(1.44e-05)
Flight FE	No	Yes	No	Yes
Observations	560,244	560,244	560,244	560,244
Log-likelihood			-52,836	-52,188

Notes: Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Demand Estimates, the role of the carrier identity

VARIABLES	(1) Pooled	(2) FE	(3) Logit	(4) Logit
LogPrice				
American	-0.00630***	-0.0116***	-0.00724***	-0.0151***
	(0.000429)	(0.000481)	(0.000480)	(0.00111)
Air Canada	-0.00573***	-0.0110***	-0.00662***	-0.00866***
	(0.000440)	(0.000491)	(0.000485)	(0.00115)
Continental	-0.00590***	-0.0113***	-0.00681***	-0.00301***
	(0.000446)	(0.000498)	(0.000495)	(0.00112)
Delta	-0.00699***	-0.0122***	-0.00828***	-0.0180***
	(0.000426)	(0.000476)	(0.000484)	(0.00189)
Lan Chile	-0.00489***	-0.0104***	-0.00593***	-0.00558***
	(0.000553)	(0.000598)	(0.000565)	(5.81e-05)
United	-0.00522***	-0.0103***	-0.00612***	-0.000838
	(0.000427)	(0.000477)	(0.000472)	(0.000832)
Advance	-0.000153***	0.000845	-0.000150***	-0.000158***
	(1.53e-05)	(0.00106)	(1.49e-05)	(1.42e-05)
Flight FE	No	Yes	No	Yes
Observations	560,244	560,244	560,244	560,244
Log-likelihood			-52,720	-52,106

Notes: Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Demand estimates, the role of the departure date

	(1)	(2)	(3)	(4)
VARIABLES	Pooled	FE	Logit	Logit
LogPrice				
December 19	-0.0113***	-0.00324***	-0.0117***	-0.0149***
	(0.000472)	(0.00108)	(0.000492)	(0.000542)
December 20	-0.00761***	-0.0163***	-0.00862***	-0.0111***
	(0.000439)	(0.00123)	(0.000459)	(0.000504)
December 21	-0.00964***	-0.0117***	-0.0102***	-0.0128***
	(0.000458)	(0.00135)	(0.000478)	(0.000525)
December 22	-0.0117***	-0.0140***	-0.0121***	-0.0154***
	(0.000476)	(0.00166)	(0.000494)	(0.000546)
December 23	-0.0108***	-0.00915***	-0.0113***	-0.0143***
	(0.000447)	(0.000832)	(0.000472)	(0.000522)
December 24	-0.0102***	-0.0189***	-0.0107***	-0.0135***
	(0.000452)	(0.00115)	(0.000474)	(0.000525)
Advance	-0.000190***	0.000740	-0.000180***	-0.000214***
	(1.55e-05)	(0.00106)	(1.47e-05)	(1.39e-05)
Flight FE	No	Yes	No	Yes
Observations	560,244	560,244	560,244	560,244
Log-likelihood			-52,476	-51,653

Notes: Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Demand estimates, the role of the departing airport

VARIABLES	(1) Pooled	(2) FE	(3) Logit	(4) Logit
LogPrice				
Newark Liberty	-0.00593***	-0.0112***	-0.00661***	-0.00665***
	(0.000436)	(0.000487)	(0.000477)	(0.000633)
La Guardia	-0.00573***	-0.0107***	-0.00639***	-0.00579***
	(0.000424)	(0.000474)	(0.000465)	(0.000563)
John F. Kennedy	-0.00679***	-0.0119***	-0.00766***	-0.0178***
	(0.000420)	(0.000471)	(0.000466)	(0.00148)
Advance	-0.000152***	0.000836	-0.000150***	-0.000160***
	(1.53e-05)	(0.00106)	(1.49e-05)	(1.43e-05)
Flight FE	No	Yes	No	Yes
Observations	560,244	560,244	560,244	560,244
Log-likelihood			-52,781	-52,157
$H_0 : \delta_{\text{EWR}} = \delta_{\text{LGA}}$ (p-value)	0.0120	1.05e-07	0.00592	0.107
$H_0 : \delta_{\text{LGA}} = \delta_{\text{JFK}}$ (p-value)	0	0	0	0
$H_0 : \delta_{\text{EWR}} = \delta_{\text{JFK}}$ (p-value)	0	6.14e-11	0	0

Notes: Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%. For k, m = EWR, LGA, JFK, and $k \neq m$ the last three rows present the p-values of the null $H_0 : \delta_k = \delta_m$.

Table 7: Demand estimates, the role of the departure time

	(1)	(2)	(3)	(4)
VARIABLES	Pooled	FE	Logit	Logit
LogPrice				
Morning	-0.00777***	-0.0140***	-0.00856***	-0.00943***
	(0.000450)	(0.000510)	(0.000495)	(0.000679)
Afternoon	-0.00736***	-0.0132***	-0.00814***	-0.00656***
	(0.000429)	(0.000485)	(0.000476)	(0.000533)
Evening	-0.00682***	-0.0126***	-0.00759***	-0.00864***
	(0.000431)	(0.000486)	(0.000474)	(0.000698)
Advance	-0.000168***	0.000714	-0.000165***	-0.000169***
	(1.54e-05)	(0.00106)	(1.50e-05)	(1.44e-05)
Flight FE	No	Yes	No	Yes
Observations	560,244	560,244	560,244	560,244
Log-likelihood			-52,791	-52,173
$H_0 : \delta_{\text{Mor}} = \delta_{\text{Aft}}$ (p-value)	7.02e-07	0	5.11e-07	6.33e-07
$H_0 : \delta_{\text{Aft}} = \delta_{\text{Eve}}$ (p-value)	1.65e-07	1.67e-10	4.00e-08	0.000688
$H_0 : \delta_{\text{Mor}} = \delta_{\text{Eve}}$ (p-value)	0	0	0	0.334

Notes: Numbers in parentheses are standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%. For k, m = Morning (Mor), Afternoon (Aft), Evening (Eve), and $k \neq m$ the last three rows present the p-values of the null $H_0 : \delta_k = \delta_m$.