View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Texas A&M University

SEARCH COSTS IN AIRLINE MARKETS

A Dissertation

by

JOSE ANTONIO PELLERANO GUZMAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Steven N. Wiggins
Co-Chair of Committee,	Steven L. Puller
Committee Members,	Timothy J. Gronberg
	Lori Taylor
Head of Department,	Timothy Gronberg

December 2013

Major Subject: Economics

Copyright 2013 José Antonio Pellerano Guzmán

ABSTRACT

This paper recovers consumer search cost estimates in airline markets consistent with theoretical search models. We follow an empirical framework developed in the recent literature on the structural estimation of search models to retrieve information on consumer search costs using price information. A unique data set of airline fares coupled with detailed ticket characteristics allows us to utilize this framework. We work in this paper with non-refundable and restricted tickets. Results show that the magnitude of search costs borne by air travelers in our ticket group is economically important. Specifically, search cost means vary between 4.75% and 8.12% of the mean route fare across markets and can be as high as US$58 for certain consumers in our sample. Consistent with previous work in other markets, our estimates indicate that most consumers sample just a few prices before buying while a relative small fraction of consumers search intensively. Results suggest that consumer search cost plays an important role in explaining part of the price dispersion observed in the airline industry.

To my parents

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Steven Wiggins, for his guidance, patience and feedback during this process. I would also like to thank my committee members, Steve Puller, Lori Taylor and specially, Timothy Gronberg, our Department Head, for his valuable insights and support in and outside the classroom. I gratefully acknowledge the financial support from the Fulbright Commission and the Department of Economics at Texas A&M University.

Thanks to my friends and colleagues who have supported me during my stay here in College Station, especially Gonzalo Sánchez, José Castillo, Jesús Bejarano, Paúl Navas, Leonardo Sánchez, Manuel Hernández, Raúl Ibarra, Manuel González and Sungmin Han.

Last but not least, I would like to thank my role models, mom and dad, for their love and support through all these years. Thanks to my brothers, Edgar and Edward. Finally, I'm truly grateful to my girlfriend Lise for her love, support and immense patience.

In memory of Prof. Sukhomlin, may the light of your inspiration enlighten us forever.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The pervasiveness of price dispersion is now well-known. Abundant theoretical and empirical literature addresses the effect of numerous factors in driving price dispersion. In particular, consumer search cost has received considerable attention from theoretical economists especially after Stigler's (1961) seminal paper "The Economics of Information". Academic interest in consumer search cost was fostered by the price dispersion observed even in markets with seemingly homogenous products. From this empirical observation it is intuitive to rationalize costly search as the underlying assumption of theoretical models since the time required to survey different sellers represents an opportunity cost to consumers. Despite the importance of consumer search behavior there are still few empirical measures of search costs in individual markets or industries.

The main goal of this paper is to contribute to our understanding of price dispersion by measuring consumer search costs in airline markets. Two distinct features of this industry suggest that consumer search costs might be important. First, consumers face a large array of fares, and second, there is considerable dispersion both within and across routes. For instance, Borenstein and Rose (1994) calculate the expected difference of fares between two passengers randomly chosen on the same route as 36 percent of the mean fare. Some within route differences can be explained by ticket characteristics associated with price discrimination or scarcity pricing. Still, part of the remaining difference is likely explained by search frictions consumers confront while booking airline tickets. In this work we recover search cost estimates of consumers in airline markets consistent with theoretical search models.

The present paper fits in the emerging literature on the structural estimation of consumer search models and builds on the work of Hong and Shum (2006). Their main insight is that the observed price distribution can be rationalized as an equilibrium outcome where profit maximizing firms respond to costly consumer search by using mixed pricing strategies. Hong and Shum show that restrictions imposed by standard search models provide sufficient structure to recover search cost estimates from pricing data alone. In our case, using only price data requires working with a group of similar tickets. Our unique dataset allows us to utilize this framework since it contains information on transacted fares coupled with detailed ticket characteristics.

This is the first paper to measure the extent of consumer search cost in the airline industry. Previous literature on airline pricing has focused on the role of market structure, peak load pricing, scarcity pricing and price discrimination strategies in driving price dispersion with little attention to consumer search cost. Our calculations suggest that search costs borne by consumers booking airline fares are economically important and that they play an important role in explaining part of the price dispersion observed in airline markets. Finally, our estimates provide further evidence that concerns about market power and policies designed to address these must consider demand side characteristics such as consumer search behavior in order to be effective.

Results indicate that mean consumer search costs in airline markets vary between \$8.84 and \$27.75 or between 4.8 and 8.1 percent of the mean fare across routes in our sample. For certain consumers, however, search costs can be as high as \$58. Furthermore, we find that most consumers sample only a few fares while a small fraction of consumers search intensively. In addition, our estimates are consistent with average markups ranging from 30.4% to 52.3% within our ticket group and routes.

It seems plausible that search costs estimates are higher offline. Our results show that consumers booking offline face search costs approximately twice as high as online consumers. A possible explanation is that online booking provides a superior search process relative to traditional offline travel agents. An alternative interpretation of this result is that incentives differ across the two distribution channels. In the case of offline booking, travel agents undertake the search process for consumers but face weaker incentives than online consumers since they are not spending (saving) their own money.

This paper is organized as follows. Section 2 reviews previous work on the emerging empirical search cost literature. Section 3 discusses the institutional set up including a succinct description of airline pricing and standard contracts in this industry. Section 4 describes the structure of the two standard search models. Section 5 describes our dataset and empirical strategy. Section 6 presents the results and Section 7 concludes.

CHAPTER II

LITERATURE REVIEW

As noted, there is an abundant theoretical literature on search costs but few empirical applications. We would like to review briefly the emerging literature on the structural estimation of consumer search models. Our analysis closely follows this previous work.

Hong and Shum (2006) build an empirical framework to retrieve consumer search cost information using only observed prices. First, they assume firms offer a homogenous product and that observed price dispersion arises due to variation in consumer search costs, caused for example, by varying opportunity costs. Given consumers' search behavior; in equilibrium, profit maximizing firms will use a mixed pricing strategy. They show that restrictions imposed by supply and demand models on the equilibrium price distribution provide sufficient structure to recover search cost estimates from observed prices.

More recently, Moraga-González and Wildenbeest (2008) develop an oligopolistic version of the non-sequential model in Hong and Shum (2006) and pose an alternative Maximum Likelihood (ML) approach for estimating this model. Monte Carlo simulations show their procedure outperforms Hong and Shum's Empirical Likelihood model both numerically and in goodness-of-fit. In this paper we use the ML procedure proposed by Moraga-González and Wildenbeest to estimate the non-sequential model.

A well cited paper from this literature is Hortacsu and Syverson (2004).[1] This is the first work to our knowledge to consider the relative role of product differentiation and consumer search cost in driving price dispersion. Using data on mutual funds and market shares they find that consumers appear to value observable non-financial features such as age, tax exposure and total number of funds in the same fund family and that even small search costs can rationalize the fact that the index fund offering the highest utility does not capture the whole market. The main difference with the present work is that we recover search cost estimates using only price data; this requires working with a group of similar tickets in our case.

Wildenbeest (2011) generalizes the model in Moraga-González and Wildenbeest (2008). In that article, the author maps a vertical differentiation model into a standard homogenous goods model which allows combining mixed strategies in prices and firm asymmetries into a single framework. Using price data from supermarkets in the United Kingdom, the structure of the equilibrium model is used to estimate both search costs and the impact of product differentiation on prices. Results indicate that around 61 percent of the variation in prices is explained by supermarket heterogeneity, while the remaining variation is due to search frictions.

One of the latest papers is De los Santos, Hortacsu and Wildenbeest (2012). They have access to a detailed dataset on the browsing and purchasing behavior of a large panel of consumers searching for books online. The availability of such a dataset allows the authors to test several predictions of the two benchmark search models. Their results

---

[1] In their model they extend the framework of Carlson and McAfee (1983).

show the non-sequential model provides a more accurate description of observed consumer search patterns. The sequential model can be rejected on both the recall patterns they observe in the data and the absence of dependence of search decisions on observed prices.

CHAPTER III

PRICING AND CONTRACTS IN AIRLINE MARKETS

Consumers booking airline tickets face a large array of fares.[2] This fare variation is one of the most salient features of the airline industry and of the complex process of airline revenue maximization. Airlines face costumers with different willingness to pay, therefore, offering different fares allows consumers to self-select themselves accordingly. On the other hand, airlines must deal with challenges specific to hospitality industries. First, inventory is highly perishable; empty seats lost value at departure. Second, airlines experience both stochastic and predictable changes in demand; load factors for specific flights are subject to random shocks and demand varies predictably along different time frames – for instance, peak hours during the day and holidays - . All these factors make revenue maximization by carriers a challenging task. Next, we discuss the two main components of airlines revenue maximization.

The first component of airlines revenue maximization is differential pricing or the practice of offering various "fare products" with different characteristics for travel on a common city-pair. These characteristics include features such as class of travel, refundability, travel and stay restrictions and advance purchase requirements.[3] Differential pricing allows carriers to screen consumers and separate them into different

---

[2] We logged on to one of the biggest online travel agencies to search for a roundtrip ticket to travel from DFW to JFK three weeks before departure. We found eighty one different fares to travel on the same day and return one week later.

[3] These different combinations of fares and ticket characteristics are commonly referred to as "bins" or "buckets" in the revenue management literature.

groups according to their price and time elasticities - e.g., business and leisure travelers – and search costs. In addition, fare departments make pricing decisions for the different groups of tickets several months in advance.

The second component of revenue maximization is yield management. Yield (or revenue) management consists in the process of determining the number of seats made available in each "fare group" previously defined by the fare department. Essentially, yield management is seat inventory control. Carriers would like to sell as many seats as possible but each sale of a discount ticket forgoes the opportunity to sell that seat at a higher price. Therefore, the main purpose of yield management is to "protect" seats for potential high-priced late booking passengers while ensuring the highest possible load factor. This process requires airlines to balance the tradeoff between capacity utilization rates and revenues per passenger.[4]

After different fare bins have been priced by the fare department and the number of tickets in each bin has been assigned by yield management, carriers make fares available through Computer Reservation Systems (CRSs). Vendors in this industry, which include traditional offline travel agents, online sellers such as Orbitz and Expedia, and airline websites, access fares through a CRS. Hence, all fares are made available through Computer Reservation Systems (CRSs).

The distribution of fares faced by travel agents is governed through a series of contracts. During our sample period, the fourth quarter of 2004, the standard contracts

---

[4] See Belobaba, Odoni and Barnhart (2009), Chapter 4, for a detailed discussion.

used by carriers included Most-Favored Nation (MFN) clauses granting agents access to all fares offered by carriers.

MFN clauses play a key role in our framework since they mean that there existed only a single price vector offered by firms which the bilateral MFN clauses require to be available both online and offline. Passengers searching for airline fares would sample from this single price vector regardless of whether they are searching online or offline. As a result, firms would recognize that the set of consumers and their associated costs would be the search costs of the entire distribution of consumers searching the entire, single price vector. This means that the search cost distribution of consumers faced by firms consists of the union of search costs of consumers searching both online and offline. For this reason, the institutional set-up requires the estimation of a single search cost distribution.

CHAPTER IV

SEARCH MODELS

The basic assumption of search models is that search is a costly activity since the time required to sample prices entails an opportunity cost for consumers. Costly search induces consumers to keep searching until the marginal cost of search equals its marginal benefit – in our set-up, the marginal benefit corresponds the expected savings of sampling an additional price - . As an abstraction of real consumers' behavior, search strategies in these models might not fit particular search habits. Still, they are useful as a general description of demand behavior when price dispersion is driven by search cost heterogeneity in the population of consumers.

The two benchmark models in this literature are the non-sequential and sequential search models.[5] The main difference between the models regards consumer search strategies. In the non-sequential model, consumers choose the number of price quotes to obtain, or alternatively, the number of times to search, and then buy at the lowest sampled price. This assumption is generally motivated by the argument that there is a fixed component of search costs. The sequential model assumes instead that consumers keep searching until the last price found is not greater than their reservation price.[6] Both models take the optimality of search strategies as given and confer little guidance on which of the two offers a better description of consumers search practices in a specific

[5] The non-sequential model is also referred to as the fixed-sample search model.
[6] Sequential search is the standard assumption in job-search models in labor economics.

empirical set-up. Some authors have maintained that a particular strategy might be preferred. For example, Manning and Morgan (1985) suggest that non-sequential search may provide a better description of the search process when there is a fixed cost component to search.

Here, we take consumer search strategies as given and retrieve search costs estimates consistent with both models. A more general model in which consumer choose their search strategy first is beyond the scope of this paper.

IV A. The Sequential Model

As discussed by Stahl (1989) most of the work in the sequential-search literature seems to have been motivated by the desire for a model in which price differences among stores could be explained as an equilibrium outcome. There was the need to fill the gap between the Bertrand and the "Diamond (1971) paradox" results. In the first case, if consumers can search costlessly (i.e., they are fully informed of the price set by each store) then the unique *Nash Equilibrium (NE)* is the competitive price. On the other hand, if search costs are non-negative, then the unique *NE* is the monopoly price. As a consequence, subsequent models featuring fully optimizing consumers and firms were developed producing mixed-strategy *NE* in prices which are interpreted as price dispersion.

Hong and Shum (2006) follow the work of Albrecht and Axell (1984), Stahl (1989) and Rob (1985) and postulate a theoretical model with heterogeneous search costs in order to generate a non-degenerate equilibrium price distribution. The intuition is that with heterogeneity in consumer search costs attributable to differences in their opportunity cost of searching, low-price firms more often serve consumers with low search costs, and high price firms more often serve consumers with high search costs. Nevertheless, as explained by the authors heterogeneity in consumer search costs is not sufficient to ensure the existence of a continuous equilibrium price distribution.

There is a set of restrictions on the population search-cost distribution that are required for non-degenerate equilibrium price dispersion in the sequential search model.[7] In particular, Theorem 4 in Rob (1985) shows that for a continuous equilibrium price distribution to exist, it must satisfy an equation like (5) below. Therefore, we can either verify that the function in (5) is a valid CDF or we can check if the likelihood function in (4) is a proper density function – that is, positive along its support – to check for the existence of a continuous price distribution.

In this model, consumer search strategy consists in choosing whether to buy at the lowest price found so far or sample an additional price, that is, there is an option value associated to searching again which resembles an "optimal stopping" problem.

There is a continuum of risk-neutral sellers producing a homogenous product at a constant marginal cost r. As discussed in Hong and Shum (2006) working only with price data does not allow for the estimation of both the distribution of search costs and marginal costs. Still, heterogeneity in marginal costs is not sufficient to generate price dispersion; high search cost consumers are required for high marginal cost vendors to

---

[7] Rob's Theorem 3 states that a sufficient condition for nonexistence of a continuous equilibrium price distribution is that the search-cost density vanishes in some positive interval $[0, A)$ with $A > 0$. In our case, we avoid this issue by assuming that search costs are distributed according to the Weibull distribution, which has support $[0, \infty)$.

sell at a high price. Each firm maximizes expected revenues by optimally choosing a price, p. Let the resulting equilibrium price distribution be $F(p)$ with bounds $\underline{p}$ and $\bar{p}$.

Similarly, the model assumes there is a continuum of buyers with differential search costs; c, randomly drawn from F_c the distribution of consumers search costs. A consumer with a cost per search of c_i will not stop searching until encountering a price $z^*(c_i)$ that satisfies the following indifference condition:

$$c_i = \int_{\underline{p}}^{z} (z-p)f(p)dp = \int_{\underline{p}}^{z} F(p)dp \qquad (1)$$

where the second equality follows from integration by parts. This indifference condition states that the consumer will stop searching when her search cost c_i equals her marginal benefit, which consists of the expected savings from making an additional search having observed z. Note that $z^*(c_i)$ is increasing in c. Define the reservation price for each c_i as:

$$p_i^* = min(z^*(c_i), \bar{p}) \qquad (2)$$

Now denote G the distribution of reservation prices in the population, given F_c and the mapping (2). Note that there is fraction of consumers with reservation price $\bar{p}$, $\alpha = 1 - G(\bar{p})$, and that $G(\underline{p}) = 0$.

The equilibrium price distribution is defined by the firms' indifference condition. Suppose consumer i has reservation price p_i^*. A firm charging $\hat{p}$ will only sell to consumers *i* for whom $\hat{p} \leq p_i^*$. Since we assume firms are symmetric, a firm's demand at price $\hat{p}$ is proportional to *(1 - G($\hat{p}$))*. Thus, the firms' indifference condition is:

$$(\overline{p} - r)D(\overline{p}) = (p - r)D(p) \Leftrightarrow (\overline{p} - r)\alpha = (p - r)(1 - G(p)) \quad (3)$$

IV B. Sequential-search Model Estimation

Equation (3) defines n-1 equations for each of the n-1 observed prices excluding $\bar{p}$. However, there are n unknowns: $G(p_1 = \underline{p}), G(p_2), \ldots, G(p_{n-1}), G(p_n = \bar{p})$. Thus, the model is under identified and additional assumptions are required for estimation. Hong and Shum (2006) propose a Maximum Likelihood estimation procedure for this model by assuming that the search cost CDF, $F_c(.;\theta)$, follows a parametric distribution family with parameters vector θ. After some algebraic manipulation, the likelihood function for each price can be derived as: [8]

$$f_p(p;\theta) = -\frac{2\alpha(\overline{p} - r)}{(p - r)^3 * f_c(c(1 - \alpha\frac{\overline{p}-r}{p-r};\theta);\theta)} - \frac{\alpha^2(\overline{p} - r)^2 f_c'(c(1 - \alpha\frac{\overline{p}-r}{p-r});\theta)}{(p - r)^4 * [f_c(c(1 - \alpha\frac{\overline{p}-r}{p-r};\theta);\theta)]^3} \quad (4)$$

With corresponding equilibrium price CDF:

$$F_p(p;\theta) = \frac{\alpha(\overline{p} - r)}{(p - r)^2 * f_c(c(1 - \alpha\frac{\overline{p}-r}{p-r};\theta);\theta)} \quad (5)$$

---

[8] Derivation of the likelihood function is presented in the Appendix.

In the above equations, we denote $c(\tau;\theta) \equiv F_c^{-1}(\tau;\theta)$, the inverse CDF of the search cost distribution. Given θ, the auxiliary parameters α and r can be solved as follows: first, the proportion of consumers with reservation price equal to $\bar{p}$, can be solved from the initial condition $G(\underline{p}) = 0$ and the indifference condition of the firm as:

$$(\overline{p} - r) * \alpha = (\underline{p} - r) * 1 \Leftrightarrow \alpha = \frac{\underline{p} - r}{\overline{p} - r} \tag{6}$$

On the other hand, selling cost r can be determined by the condition that $F_p(p) = 1$. Therefore, given θ and equations (5) and (6), r must satisfy:

$$1 = F_p(\overline{p}) = \frac{(\underline{p} - r)}{(\overline{p} - r)^2 * f_c(c(1 - \frac{\underline{p}-r}{\overline{p}-r};\theta);\theta)} \tag{7}$$

The likelihood function for the whole sample of prices, then, is $L(\theta, r) = \prod_i f_p\,(p_i;\theta)$.

As discussed above, a necessary condition for the existence of a non-degenerate continuous equilibrium price distribution for this model within an interval $[\underline{p},\bar{p}]$ is that the price CDF in (5) is nondecreasing in this range, or equivalently, that the likelihood function (4) be positive for all $p \in [\underline{p},\bar{p}]$. This is a strong condition since it requires after examining equation (4) that the search cost density f_c to be strictly decreasing in the range $[0, c(\frac{\bar{p} - p}{\bar{p} - r}); \theta)]$.

In this paper, we use the Weibull distribution for the search cost distribution. Besides its flexible shape the Weibull distribution is one of the few two-parameter distributions that allows for the density to be strictly decreasing along its full support. For the Weibull density,

$$f(x|\theta_1,\theta_2) = \frac{\theta_2}{\theta_1}\left(\frac{x}{\theta_1}\right)^{\theta_2-1} e^{-(x/\theta_1)^{\theta_2}}, \qquad \theta_1,\theta_2 > 0$$

the decreasing density condition restricts $\theta_2 \leq 1$. This restriction has important implications for the results as we discussed further in the next section.

IV C. The Non-sequential Model

The following discussion is based on Moraga-González and Wildenbeest (2008). They study an oligopolistic version of the model proposed in Hong and Shum (2006). Hong and Shum generalize the non-sequential search model of Burdett and Judd (1983) by adding search cost heterogeneity.

The basic structure of the model is as follows. There are *N* retailers selling a homogenous good with a common unit selling cost *r*. There is a unit mass of buyers who differ in their search costs which are randomly drawn from F_c, the consumers search cost distribution. Consumers have inelastic demand and buy one unit of the product. Let $\bar{p}$ be the consumer valuation for the good. A standard assumption of search models is that consumers obtain a first price quote at random for free. Beyond the first price, a consumer incurs a search cost *c* per additional price quote. Therefore, consumer with search cost *c* incurs in total search cost $c*i$.

Following the notation in Burdett and Judd (1983), denote the symmetric mixed strategy equilibrium by the distribution of prices F_p, with density $f_p(p)$. Let $\underline{p}$ and $\bar{p}$ be the lower and upper bound of the support of F_p. Given the behavior of firms, consumer with search cost c will choose optimally the number of prices $i(c)$ she observes in order to minimize total search costs $c*(i-1)$, i.e.:

$$i(c) = arg\min_{i>1} c(i-1) + \int_{\underline{p}}^{\overline{p}} ip(1-F_p(p))^{i-1} f_p \, dp \qquad (8)$$

since $i(c)$ must be an integer, this search strategy splits consumers into N subsets of size q_i, $i = 1, 2, \ldots, N$, with $\sum_{i=1}^{N} q_i = 1$; then, q_i is the fraction of buyers sampling i stores (prices) and is strictly positive for all i. This partition is computed as follows. For the consumer indifferent between sampling i and $i+1$ firms it must hold that her search cost equals:

$$\Delta_i = Ep_{1:i} - Ep_{1:i+1}, i = 1, 2, ..., N-1 \qquad (9)$$

here $E_{p1:i}$ denotes the expected minimum price in a sample of i prices drawn from the price distribution F_p. Note that i is decreasing in c. Using this property, the fractions of consumers q_i sampling i prices are then:

$$\begin{aligned} q_1 &= 1 - F_c(\Delta_1), i = 1, 2, ..., N-1; \\ q_i &= F_c(\Delta_{i-1}) - F_c(\Delta_i), i = 2, 3, ..., N-1; \\ q_N &= F_c(\Delta_{N-1}). \end{aligned} \qquad (10)$$

Figure 1 presents a graphical illustration. Assume this graph depicts the recovered search cost distribution; q_1 and Δ_1 represent the fraction of consumers who do not incur in any search and their corresponding search cost, respectively. Similarly, q_2 is the fraction of consumers who survey at most two prices before buying and Δ_2 denotes their corresponding search cost. Note from the graph that Δ_i should be interpreted as a lower bound for the search cost since it coincides with the search cost of the marginal consumer who is just indifferent between searching *i* prices and *i+1* prices. Also notice that we can only identify the shape of the search-cost distribution up to the $1 - F(\Delta_1)$ percentile.

Turning to the firms' problem, the upper bound of the price distribution must be $\bar{p}$ since a firm charging this price will only sell to consumers who do not incur in any search, i.e., consumers in q_1, who do not compare prices and accept to buy at $\bar{p}$. For the equilibrium price distribution to represent a mixed strategy equilibrium, firms must be indifferent between charging the highest price, $\bar{p}$, and any other price in the support of F_p. Therefore, the equilibrium condition is:

$$(p - r)[\sum_{i=1}^{N} \frac{iq_i}{N}(1 - F_p(p))^{i-1}] = \frac{(\bar{p} - r)q_1}{N}, \ \forall p \ \epsilon \ [\underline{p}, \bar{p}] \quad (11)$$

From this equilibrium condition, it follows that the minimum price charged in the market is:

$$\underline{p} = \frac{(\bar{p} - r)q_1}{\sum_{i=1}^{N} iq_i} + r \quad (12)$$

Hong and Shum (2006) show that equations (8) to (12) provide enough structure to retrieve information on search costs using only price data for the non-sequential model.

IV D. Non-sequential Model Estimation

Moraga-González and Wildenbeest (2008) proposed to estimate the $\{\Delta_i, q_i\}$ for $i = 1, \ldots, N$ pairs of the search cost distribution by Maximum Likelihood as follows since equation (11) cannot be solved analytically for the equilibrium price distribution F_p which makes it difficult to calculate the cutoff points:

$$\Delta_i = \int_{\underline{p}}^{\overline{p}} p[(i+1)F_p(p) - 1](1 - F_p(p))^{i-1} f_p(p)dp, i = 1, 2, ..., N-1.$$

In their work Hong and Shum (2006) propose to use the observed price distribution to calculate the Δ_i's. Moraga-González and Wildenbeest (2008) through Monte Carlo simulations show that, although practical, this procedure does not necessarily provide minimal variance estimates. Therefore, they propose an alternative method to obtain ML estimates of the search cost distribution cutoff points. To do this, we have to rewrite first Δ_i as a function of the ML estimates of the parameters of the price distribution. Integrating by parts, we first rewrite the cutoff points as:

$$\Delta_i = \int_{\underline{p}}^{\overline{p}} F_p(p)(1 - F_p(p))^i dp, i = 1, 2, ..., N-1. \qquad (13)$$

Since F_p is monotonically increasing in p, its inverse exists so we can use equation (11) to solve for p as:

$$p(z) = \frac{(\overline{p} - r)q_1}{\sum_{i=1}^{N} iq_i(1-z)^{i-1}} + r \qquad (14)$$

Using (14) and a change of variables equation (9) can be written as:

$$\Delta_i = \int_0^1 p(z)[(i+1)z - 1](1-z)^{i-1}\, dz,\ \ i = 1,..,N-1 \quad (15)$$

Therefore, given ML estimates of r, $\underline{p}$, $\bar{p}$, and q_i, $i = 1, 2,...,N$, we can use equations (14) and (15) to obtain ML estimates of the cut-off points of the search cost distribution by the invariance property of ML estimation.

Now, to obtain estimates of r, $\underline{p}$, $\bar{p}$, and q_i, $i = 1, 2,...,N$, by ML using only price data we first apply the implicit function theorem to equation (11), which yields:

$$f_p(p) = \frac{\sum_{i=1}^{N} iq_i(1-F(p))^{i-1}}{(p-r)\sum_{i=1}^{N} i(i-1)q_i(1-F_p(p))^{i-2}} \quad (16)$$

Let's order the vector of observed prices in ascending order $p_1, p_2, ..., p_M$. We use p_1 and p_M as estimates of the lower and upper bounds of the support of the price distribution $\underline{p}$ and $\bar{p}$, respectively.[9] Using these, we can solve for the marginal cost as a function of the other parameters using equation (12) above:

$$r = \frac{\sum_{i=1}^{N} iq_i - q_1 p_m}{\sum_{i=2}^{N} iq_i} \quad (17)$$

---

[9] $\underline{p}$ and $\bar{p}$ converge super consistently to the true bound of the price distribution. See Donald and Paarsch (1993) for further discussion.

Finally, plugging r into the density function in (16) and using the fact that $q_N = 1 - \sum_{i=1}^{N-1} q_i$ the next maximum likelihood problem can be solved numerically:

$$\max_{q_i{}_{i=1}^{N-1}} = \sum_{l=2}^{M-1} log\ f_p(p_l; q_1, q_2, ..., q_N) \tag{18}$$

where

$$F_{p_l}\ solves\ \frac{(\overline{p_l} - r)q_1}{N} = (p - r)[\sum_{i=1}^{N} \frac{iq_i}{N}(1 - F_{p_l}(p))^{i-1}]\ \forall l = 1, ..., M-1$$

The standard errors of the estimates of q_i, $i = 1, 2,...,N-1$ are calculated by taking the square root of the diagonal entries of the inverse of the negative Hessian matrix evaluated at the optimum as usual. Standard errors of q_N, r and Δ_i*'s* can be computed by the Delta method.

The estimation procedure work as follows. Given arbitrary starting values $\{q_i^0\}_{i=1}^{N-1}$. Then for every price p_l in the data set, $F_p(p_l)$ is calculated using the equilibrium condition (11), which in turn allows us to calculate $f_p(p_l)$ using (16). A Preconditioned Conjugate Gradient (PCG) method is then used which change the $q_i's$ until the log-likelihood function is maximized.

CHAPTER V

EMPIRICAL APPLICATION

V A. Data and Empirical Strategy

Our dataset contains all online and offline ticket transactions made through one of the major CRS for the fourth quarter of 2004 and represents roughly thirty percent of all domestic tickets sold in the U.S. during that period. Fares booked to travel after December 22nd through the end of year are excluded as well as those booked the Wednesday before Thanksgiving through the following Sunday. Data spans transactions from offline travel agents, airline websites and several online booking sites. This first database includes airline and flight number, origin and destination, fare, booking class, a fare code, and dates of purchase, departure and return but does not include information on ticket characteristics.

Ticket characteristics were retrieved from another CRS archive containing both fares offered and purchased for travel in particular city-pairs organized by departure, airline, and city-pair. From this second data set we collect information on carrier, origin and destination, departure date, fare, booking class, advance purchase requirements, refundability, travel restrictions, and minimum and maximum stay restrictions for each fare. These data were then matched to the transactions database. The criterion used in the matching process was to keep a transaction if this could be matched to a fare in the second dataset within two percent; for multiple matches within two percent the closest was kept. Additional details of the matching process are presented in Sengupta and

Wiggins (2013) where they also present evidence that this procedure did not introduce major selection issues. The availability of such a dataset allows us to utilize the theoretical framework presented in the previous section since retrieving search cost information using only price data requires a sample of consumers booking a group of similar tickets.

Airline tickets can be grouped in a price ascending order in four main categories: non-refundable and restricted fares, non-refundable and non-restricted fares, refundable and non-restricted fares and first class. We work in this paper with non-refundable and restricted tickets. In general, this is the cheapest category of fares and the typical ticket booked by leisure travelers, the most price sensitive and time insensitive group of consumers. Leisure or vacation travelers are willing to meet basically any travel and ticket conditions in order to pay the lowest fare. Given these characteristics, our model's assumption of symmetric firms serving consumers who search for the lowest fare seems appropriate. The latter becomes more relevant in the actual setup as we do not consider substitution across carriers. Worries about carriers' choice and consumer loyalty should be more of a concern for higher priced ticket groups, especially, first class and refundable and non-restricted fares.[10]

Our dataset contains information on 250 domestic routes defined at the airport-pair level regardless of direction. The routes in this paper were selected based on the number of observations within our ticket group in the transactions dataset from the first CRS.

---

[10] Following Wildenbeest (2011), it can be argued that the fraction of consumers searching only once can be interpreted as carrier loyal customers; therefore, up to some extent loyalty can be accommodated in our setting.

The prices used are for roundtrip fares only excluding itineraries with open-jaws and circular trip tickets and include tickets for flights operated by American Airlines, Continental, Delta, Northwest, US Airways, United Airlines, Frontier, Air Tran, Spirit, Alaska, America West, Sun Country, Hawaiian Airlines and American Trans Air.

Table 1 presents summary statistics of fares in our ticket group for the four routes included in this work: ATL-ORD, DTW-LGA, ORD-PHX and ORD-SFO.[11] The first row presents summary statistics for all observations in these routes, and the next two rows, for offline and online transactions separately. The information on this table suggests that consumers booking airline tickets in these markets have incentives to search given the potential savings from sampling additional fares. Standard deviations range from 15% to 30% of the mean route fare across routes. Additionally, summary statistics indicate that online fares present lower means than offline fares and less dispersion in general.

Within our ticket group, non-refundable and restricted fares, we limit the sample to tickets booked eight to fourteen days before departure. This lessens the concerns of availability and potential price changes by airlines as a response to unexpected increases in bookings occurring during the last seven days prior departure. Table 2 presents summary statistics of fares in our ticket group for the 8-10 day window and the 11-14 day window before departure. As you can see from the table summary statistics are very close across the two windows; mean and median fares in the 8-10 day window are not

[11] ATL: Hartsfield-Jackson Atlanta International Airport; DTW: Detroit Metropolitan Wayne County Airport; LGA: La Guardia Airport; ORD: Chicago O'Hare International Airport; PHX: Phoenix Sky Harbor International Airport; SFO: San Francisco International Airport.

more than 3 percent higher than mean and median fares for the second window. Only for the ORD-SFO route the 8-10 day window mean fare is approximately 5 percent higher although this difference is not statistically significant.

Previous papers in general use information on posted prices. This raises two potential concerns. First, posted prices might bias upwards search cost estimates if high price quotes do not generate any trade; and second, “bait and switch” practices in which firms post low prices just to attract potential consumers could induce a downward bias in recovered search costs. We avoid these issues by working with ticket transactions.

We are using the distribution of observed transactions during the second week before departure as a proxy for the unknown price distribution at every point in time during the second week before departure in our sample period. Therefore, we are assuming the same fares are available throughout this period. As discussed in section 3, fare departments set prices for the different ticket groups far in advance the flying date, and also, yield management implies that some fares might become unavailable at some point before departure. This might generate an upward bias in search cost estimates since we do not observe stock outs. Still, the fact that we consistently observe in the dataset consumers booking fares in the cheapest group even up to the last day before departure is evidence that these fares are available through the whole sample period.

CHAPTER VI

ESTIMATION RESULTS

VI A. Sequential Model Results

In this section we present results for the sequential search model. The equilibrium condition of firms in the sequential model defines *n-1* equations, one for each of the *n* observed prices excluding the highest price.[12] However, there are *n* unknowns; the *n* percentiles of the reservation price distribution corresponding to each of the *n* observed prices. As a consequence, additional assumptions are required to estimate this model. Hong and Shum (2006) propose to estimate this model by MLE assuming the search cost distribution F_c is parameterized by a finite vector θ. Note that by assuming a functional form for the search cost distribution we can extrapolate the entire search cost distribution.

As just mentioned, ML estimation of the sequential model implies specifying a functional form for the search costs pdf. Moreover, model derivation requires the assumed function to be decreasing along all of its support in order for the density of prices to be a valid pdf.[13] Only a few distribution families satisfy this restriction. For this work we explored different specifications using the Weibull, the gamma and the log-normal distributions. Results were similar across the different specifications.

---

[12] See equation 3 in section 4.
[13] See equation 4.

Table 3 presents results for the four routes using the Weibull distribution. Columns 1 and 2 present the estimates of the shape and scale parameter respectively. The table also shows results for the mean and the median of search costs, the fraction of consumers who do not incur in any search and their corresponding search cost estimate. Note the high search cost means and the large search cost estimates for α, the fraction of consumers who do not search. Results show that search costs borne by consumers in this group start at US$255.5 in the ORD-SFO route. In three of the four cases these search cost estimates are above the mean route fare.

As discussed in Hong and Shum (2006), these high sequential search cost estimates can be explained by the behavior of the parametric families we consider. The restriction imposed on the shape parameter of the parametric distribution in order for the slope of the search cost density to be negative along all of its support makes this density to die-off very slowly.[14]

Recovered estimates from this model exceed reasonable expectations for consumers search costs in these markets. For this reason, discussion of results in this paper is based on non-sequential search costs estimates.

VI B. Non-Sequential Model Results

This section presents the estimates of the non-sequential model for the four routes in our sample. The estimates from this model are search cost distribution pairs $\{q_i, \Delta_i\}$ for $i = 1, 2, \ldots, N$, these are, the fractions of consumers who search i prices and their

[14] The Weibull distribution is decreasing in all its support conditional on the shape parameter being less or equal to 1.

corresponding search cost estimate in dollars. Recall that Δ_i's should be interpreted as lower bounds - consumers who search i prices must realize search costs of at least Δ_i - and also that we can only identify the shape of the search-cost distribution up to the $1 - F(\Delta_1)$ percentile.

To estimate this model we use the MLE procedure in Moraga-González and Wildenbeest (2008). One of the insights of their estimation procedure is that once we have obtained ML estimates of the q_i's, that is, the fractions of consumers searching i prices, for $i = 1, 2,\ldots,N$, we can obtain corresponding ML of the search cost estimates, Δ_i's, by first rewriting prices as a function of the q_i's.[15] We can then compute standard errors using the Delta method as usual.

Table 4 presents the results for the non-sequential model. Columns 2 through 4 contain the ML estimates for each route: p_1 and v correspond to the route lowest and highest observed fare, respectively; N is the number of unique fares rounded to the closest integer and M is the number of observations for the sample period in our ticket group. Next we present ML estimates for the fraction of consumers searching i prices: q_1 is the fraction of consumers that do not incur in any search, q_2 is the fraction of consumers searching at most two prices and so forth, up to q_n, which represents the fraction of consumers who search exhaustively. Additionally, the table presents the estimate for the marginal cost r, the value of the log-likelihood function and the computed value for the Kolmogorov-Smirnov statistic which test the null hypothesis that the vector of estimated and observed fares belong to the same continuous distribution.

[15] See equations 14 and 15 in section 4.

Panel B below shows the corresponding ML estimates for the search cost values for each q_i. The numbers in parenthesis correspond to the standard errors of estimates.

Results show that consumers incur in little search overall. The estimate for q_1 is 0.25, 0.21, 0.21 and 0.35 for each of the four routes, respectively. In the case of ATL-ORD, the estimate for q_2 is 0.66 which means that 91 percent of consumers search at most two prices before buying on this route. Doing a similar exercise across the other three routes the reader can verify that 81 and 90 percent of consumers search at most seven prices on the DTW-LGA and ORD-PHX routes, respectively, and that approximately 90 percent of consumers search four prices at a maximum on the ORD-SFO route. Air travelers who search intensively account for less than 10 percent, excluding DTW-LGA, where this fraction is 19 percent. Estimates are statistically significant in general and all other q_i's are not statistically different from zero. This result is consistent with previous findings in other markets; most consumers incur in little search while a small proportion of consumers search intensively.

As can been seen from panel B, consumers who do not search must bear search costs of at least US$ 18.38 on the ATL-ORD route, whereas, on the ORD-SFO route this number is at least US$ 58.35. Alternatively, expected savings from sampling at least two prices are US$ 18.38 on the ATL-ORD route and US$ 58.35 on the ORD-SFO route. On the other hand, consumers who are indifferent between sampling two or three prices realize search costs of US$ 6.43 in the first route and US$ 28.08 on the fourth route – that is, expected savings from adding a third price to the search vary between US$ 6.43 and US$ 28.08 across the four routes - . All estimates are highly significant. The last row in panel B shows that the route weighted mean search cost ranges from US$ 8.84 in the ATL-ORD route up to US$ 27.75 in the ORD-SFO route, or, 4.8 and 8.1 percent of the respective mean route fare.

We also recover estimates of the marginal cost *r*; all estimates are within 6 percent of the lowest observed fare. These marginal cost estimates are consistent with average markups across routes ranging from 30.4 to 52.3 percent. This result is notable considering we are working with a narrowly defined ticket group and characteristics such as class of travel, refundability, and travel restrictions have already been considered.

Finally, we compare consumers search costs offline and online, ex-post we first calculate the fares corresponding to each search cost cutoff point and then obtain the percentages of transactions between these fares offline and online separately. Next, we compute a weighted search cost by multiplying these percentages by the corresponding search cost cutoff point. Again, these should be interpreted as lower bound estimates for search costs online and offline. Table 5 presents the results. Offline consumers bear search costs at least 71 percent higher than online consumers across these markets. In all four routes offline search costs nearly double online estimates as a percentage of the mean route fare. An interpretation of this result could be that online booking offers a less costly search mechanism relative to offline travel agents. It is also plausible that this difference is explained by different incentives across the two distribution channels. Offline travel agents making the search for consumers face weaker incentives to search since they are not spending (saving) their own money, while online consumers, spending their own money, face greater incentives to search.

CHAPTER VII

SUMMARY AND CONCLUSIONS

This paper recover search cost estimates of consumers booking airline fares. Results show that the magnitude of search costs borne by consumers in our ticket group is economically important specially if we consider that this ticket group is usually booked by leisure travelers, arguably the segment of air travelers with lower search costs in general. Specifically, average search costs estimates range between \$8.84 and \$27.75 across the four routes in our sample – or between 4.8 and 8.1 percent of the route mean fare - and can be at least as high as \$58.4 for consumers who do not search in one of the markets examined in this paper. With respect to the amount of search consumers engaged in, our results are consistent with previous findings in other markets; most consumers incur in little search and a small fraction of consumers search intensively. Additionally, recovered marginal costs allow us to compute estimates of average mark-ups within routes. These estimates vary between 30.4 and 52.3 percent across routes included in this work.

Results also indicate that consumers booking online realize lower search costs than consumers booking via traditional offline agents. This result seems intuitive as the internet might offer a less costly search process to consumers booking online, and also as agents face different incentives across the two distribution channels. Consumers booking online have higher incentives to search and book the cheapest available fares since they are paying for the ticket, whereas, offline travel agents undertaking the search process for consumers face weaker incentives to search as much since they are not spending their own money. However, these results should be interpreted cautiously. The choice of distribution channel is an endogenous process so we cannot readily interpret that differences in search costs online and offline arise from inherently higher or lower costs for either channel. This result might also be explained because of a combination of opportunity costs, internet familiarity and other differences between the two groups of consumers.

This paper contributes to the emerging empirical literature estimating search cost estimates consistent with theoretical search models. This is the first paper to recover consumer search cost estimates in the airline markets. Regarding the literature on airline pricing, extensive theoretical and empirical work has addressed the roles of market structure, peak-load pricing, scarcity pricing theories and price discrimination strategies in driving price dispersion. So far, little attention has been devoted to consumer search behavior in these markets. Our results suggest that consumer search cost plays an important role in driving part of the price dispersion observed in this industry.

In general, policies designed to address market power concerns have focused on supply side factors. Results from this recent literature including this work suggest that demand factors such as consumer search behavior are also of practical importance.

REFERENCES

Albrecht, James and Bo Axell (1984), “An Equilibrium Model of Search Unemployment", *Journal of Political Economy*, *Vol. 92*, pp. 824-40.

Belobaba, Peter, Amadeo Odoni and Cynthia Barnhart (2009), The Global Airline Industry (Chippenham, Wiltshire, United Kingdom: John Wiley & Sons, Ltd., 2009).

Borenstein, Severerin and Nancy Rose (1994), “Competition and Price Dispersion in the U.S. Airline Industry", *Journal of Political Economy*, *Vol. 102 (4)*, pp. 653-83.

Burdett, Kenneth and Kenneth L. Judd (1983), “Equilibrium Price Dispersion", *Econometrica*, *Vol. 51*, pp. 955-69.

Carlson, John A. and R. Preston McAfee (1983), “Discrete Equilibrium Price Dispersion”, *Journal of Political Economy*, *Vol. 91 (3)*, pp. 480-93.

Diamond, Peter (1971), “A Model of Price Adjustment", *Journal of Economic Theory*, *Vol. 2*, pp. 156-68.

Donald, Stephen J. and Harry J. Paarsch (1993), “Piecewise Pseudo-Maximum Likelihood Estimation in Empirical Models of Auctions", *Econometrica*, *Vol. 34*, pp. 121-48.

Hong, Han and Matthew Shum (2006), “Using Price Distributions to Estimate Search Costs", *RAND Journal of Economics*, *Vol. 37 (2)*, pp. 257-75.

Hortacsu, Ali and Chad Syverson (2004), “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: a Case Study of S & P 500 Index Funds", The *Quarterly Journal of Economics*, *Vol. 119*, pp. 403-56.

Moraga-Gonzalez, Jose L. and Matthijs R. Wildenbeest (2008), “Maximum Likelihood Estimation of Search Costs", *European Economic Review*, *Vol. 52*, pp. 820-48.

Morgan, Peter and Richard Manning (1985), “Optimal Search", *Econometrica*, *Vol. 53*, pp. 923-44.

Rob, Rafael (1985), “Equilibrium Price Distributions", *Review of Economic Studies*, *Vol. 52*, pp. 487-504.

Sengupta, Anirban and Steven Wiggins (2013), "Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet", *American Economic Journal: Economic Policy*, forthcoming.

Stahl, Dale O. (1989), "Oligopolostic Pricing with Sequential Consumer Search", *American Economic Review*, *Vol. 79*, pp. 700-12.

Stigler, George J. (1964), "The Economics of Information", *Journal of Political Economy*, *Vol. 61*, pp. 213-25.

Wildenbeest, Matthijs R. (2011), "An Empirical Model of Search with Vertically Differentiated Products", *RAND Journal of Economics*, *Vol. 42(4)*, pp. 729-57.

Supplemental Sources Consulted

Dana, James D., Jr. (1998), "Advance-purchase Discounts and Price Discrimination in Competitive Markets", *Journal of Political Economy*, *Vol. 101 (2)*, pp. 395-422.

________ (1999), "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure", *RAND Journal of Economics*, *Vol. 30*, pp. 632-60.

Gale, Ian and Thomas Holmes (1993), "Advance-Purchase Discounts and Monopoly Allocation of Capacity", *American Economic Review*, *Vol. 83*, pp. 135-46.

Gerardi, Kristopher S. and Adam H. Shapiro (2009), "Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry", *Journal of Political Economy*, *Vol. 117 (1)*, pp. 1-37.

Hernandez, Manuel and Steven Wiggins (2013), "Nonlinear Pricing Strategies and Competitive Conditions in the Airline Industry", *Economic Inquiry*, forthcoming.

Puller, Steven, Anirban Sengupta, and Steven Wiggins (2012), "Does Scarcity Drive Intra-route Price Dispersion in Airlines?", Working paper Texas A&M University.

Stavins, Joanna (2001), "Price Discrimination in the Airline Market: The Effect of Market Concentration", *Review of Economics and Statistics*, *Vol. 83 (1)*, pp. 200-02.

APPENDIX A

FIGURES AND TABLES

Figure 1 - Identification Scheme for Search-Cost Distribution in Non-Sequential Search Model

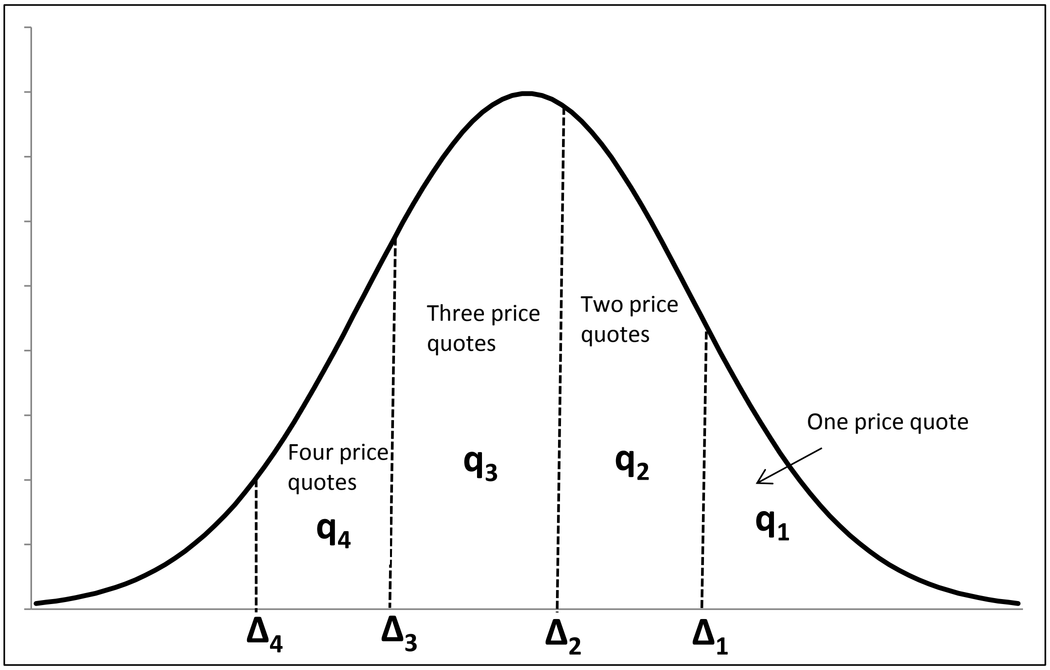

Figure 2 – Fares Distribution ATL-ORD

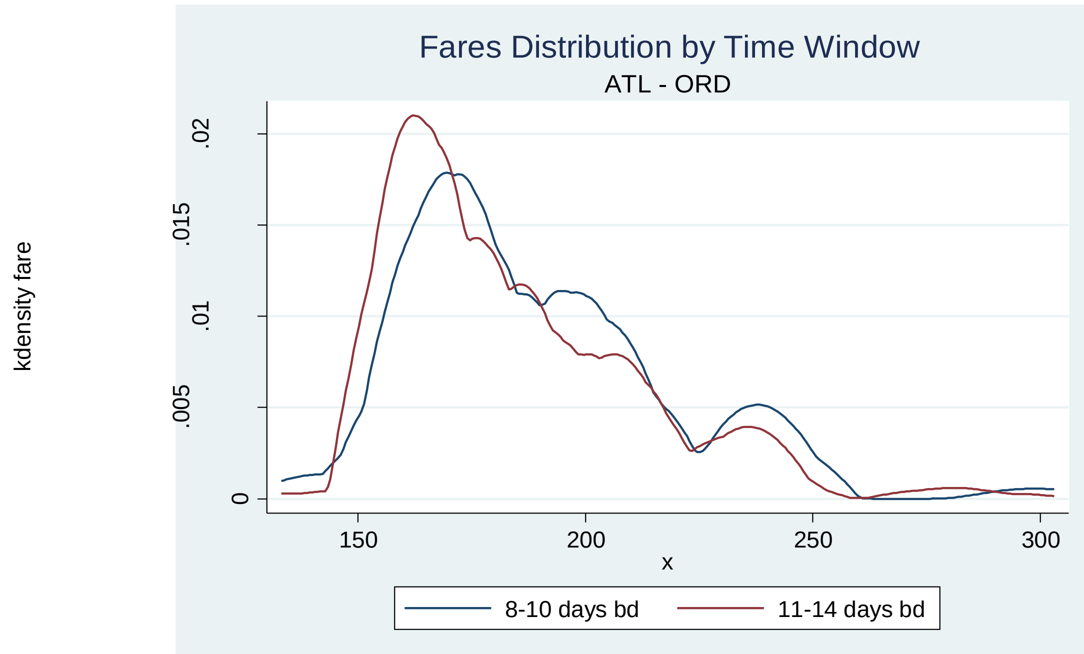

Figure 3 - Fares Distribution DTW-LGA

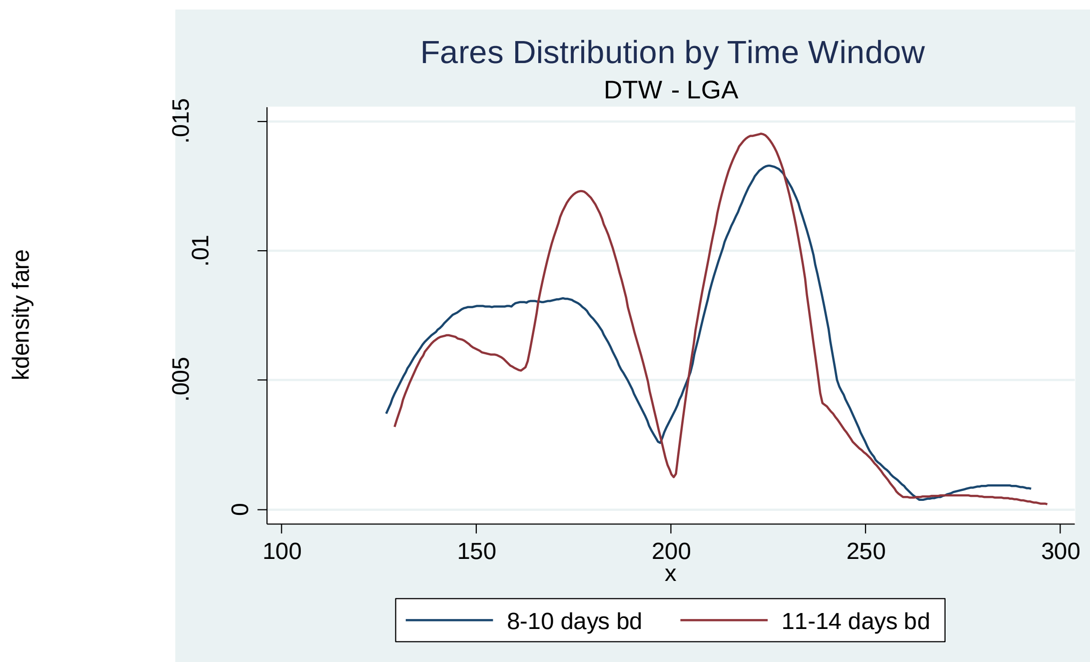

Figure 4 - Fares Distribution ORD-PHX

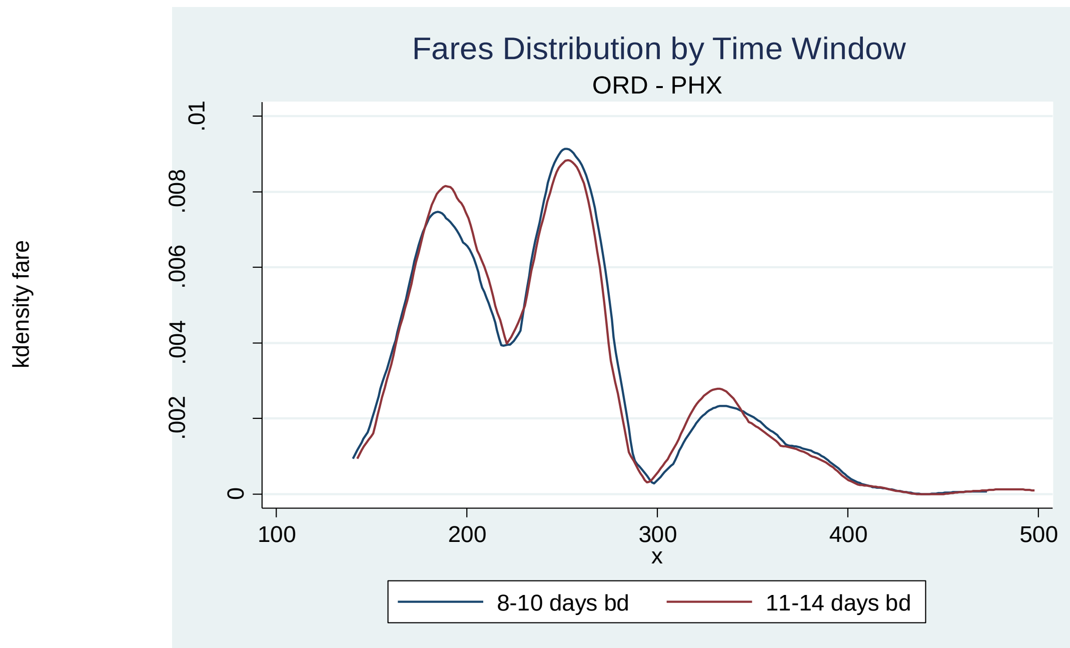

Figure 5 - Fares Distribution ORD-SFO

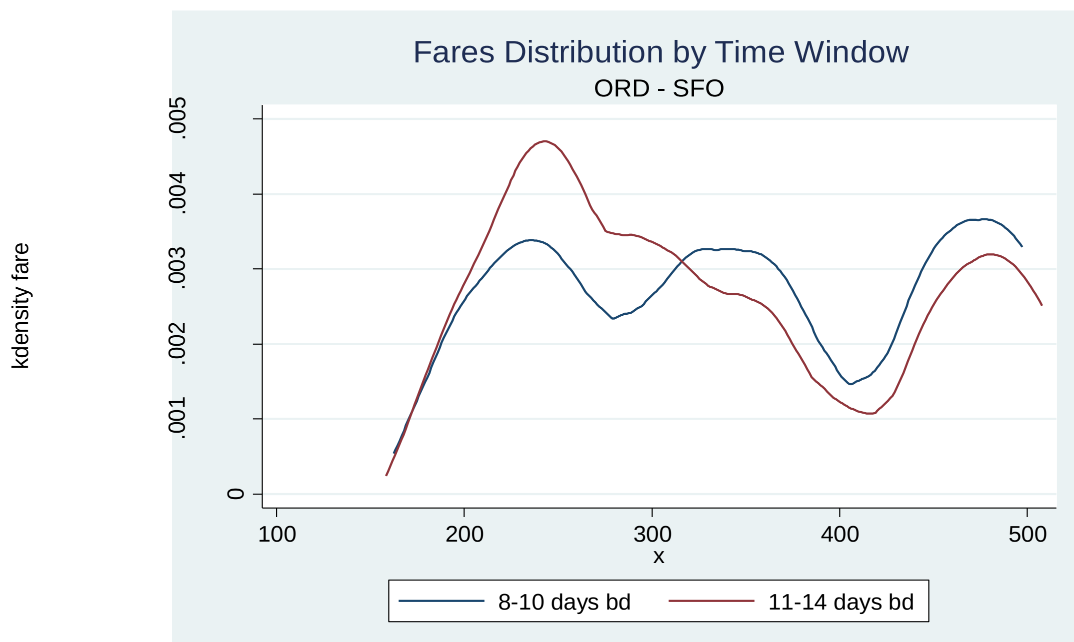

Figure 6 - ATL-ORD Sequential Estimation

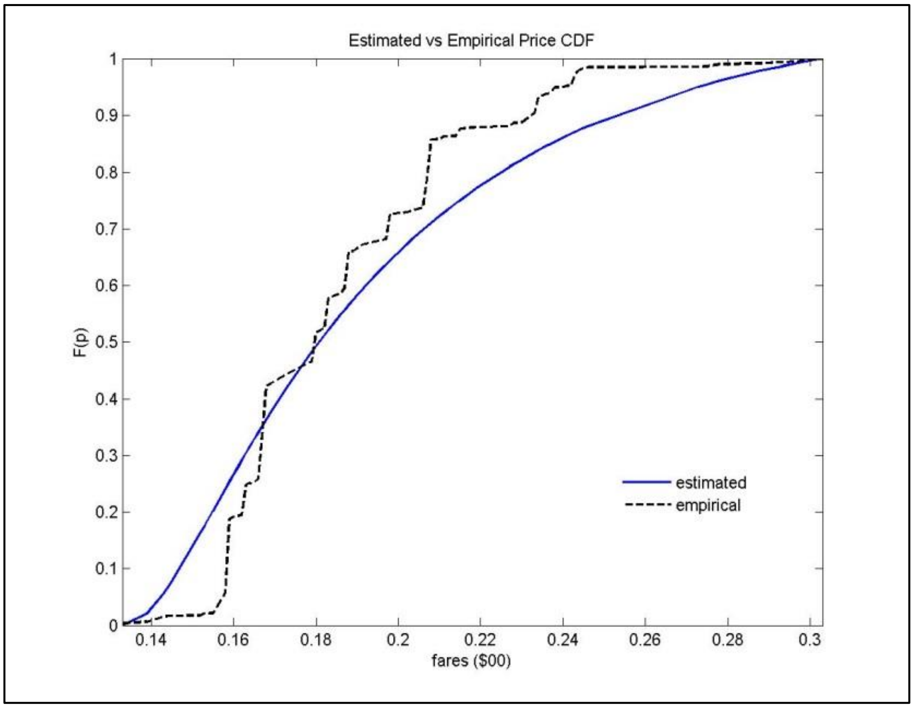

Figure 7 - DTW-LGA Sequential Estimation

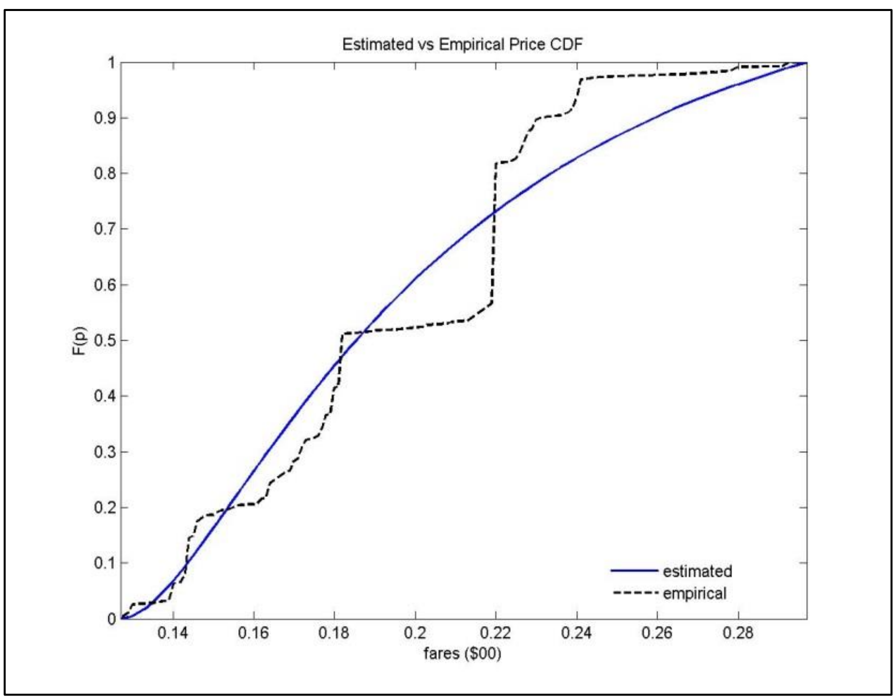

Figure 8 - ORD-PHX Sequential Estimation

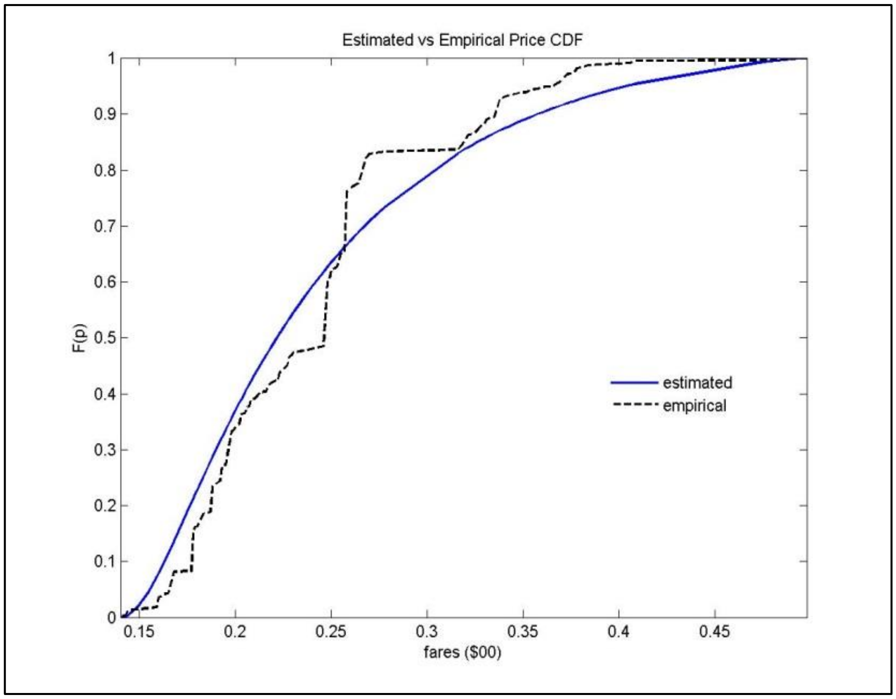

Figure 9 - ORD-SFO Sequential Estimation

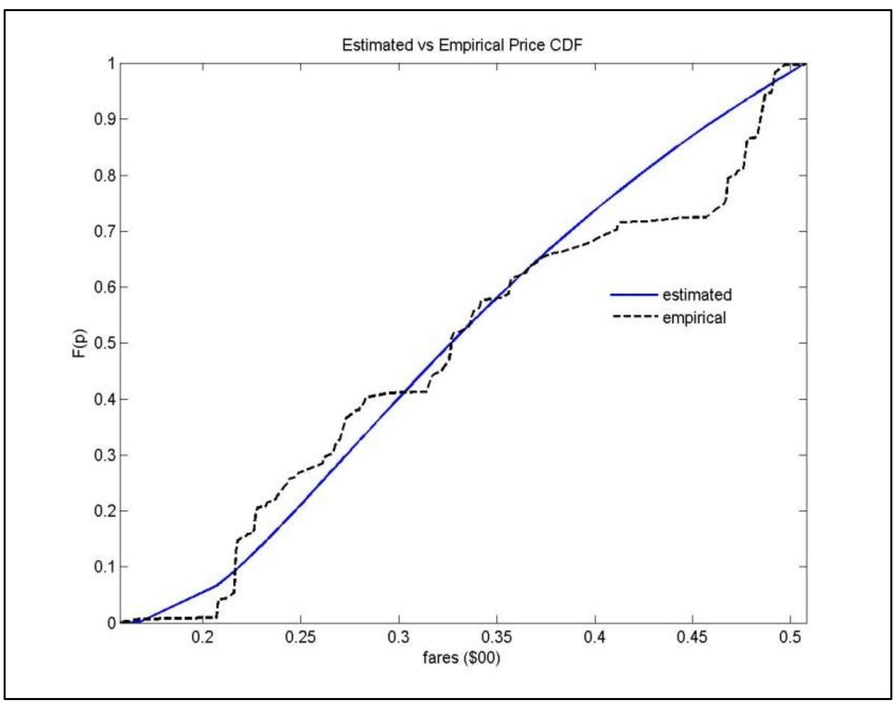

Figure 10 - ATL–ORD Non Sequential Estimation

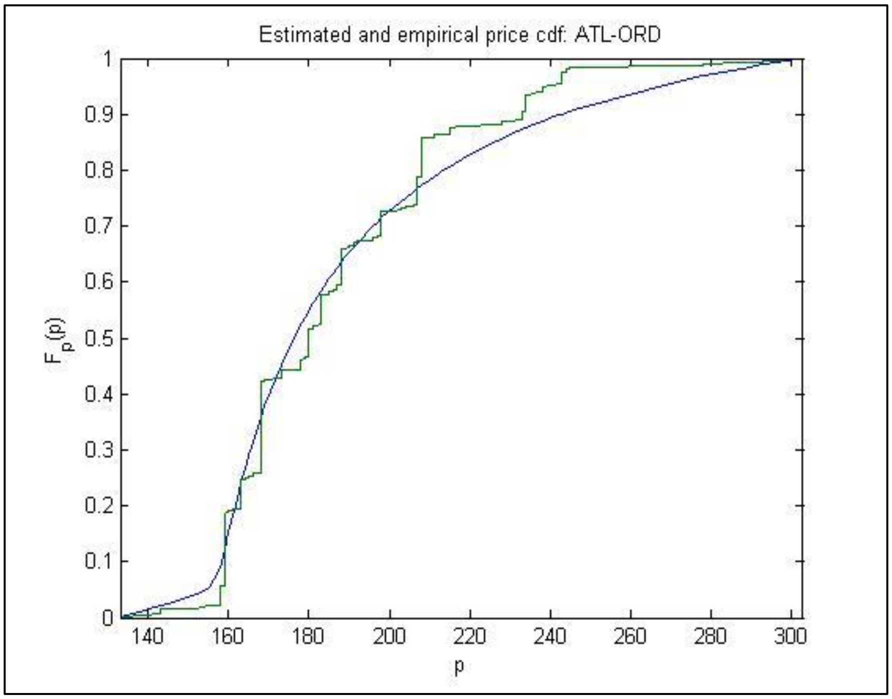

Figure 11 - DTW-LGA Non Sequential Estimation

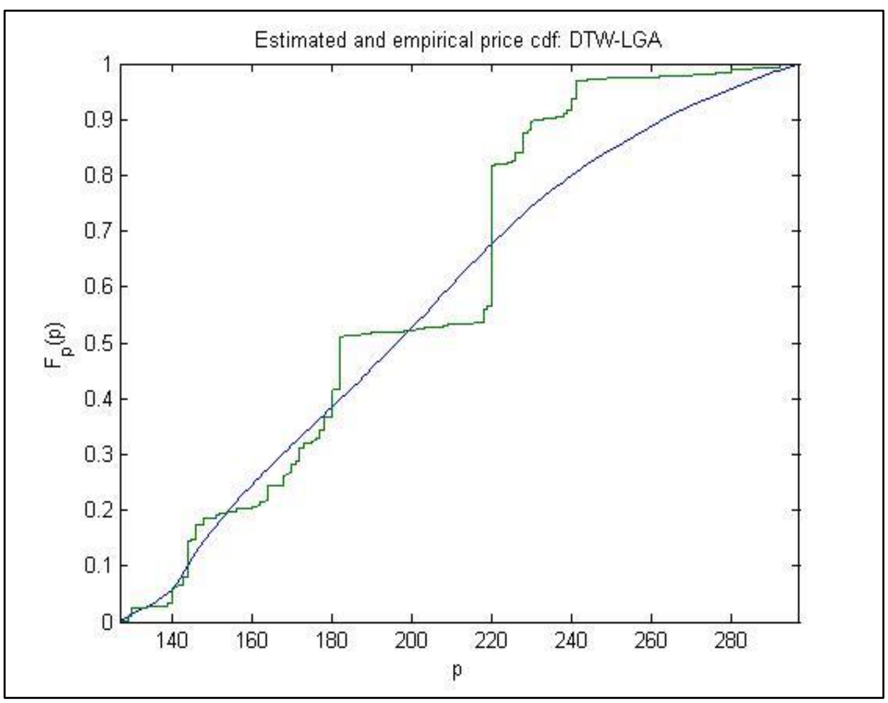

Figure 12 - ORD-PHX Non Sequential Estimation

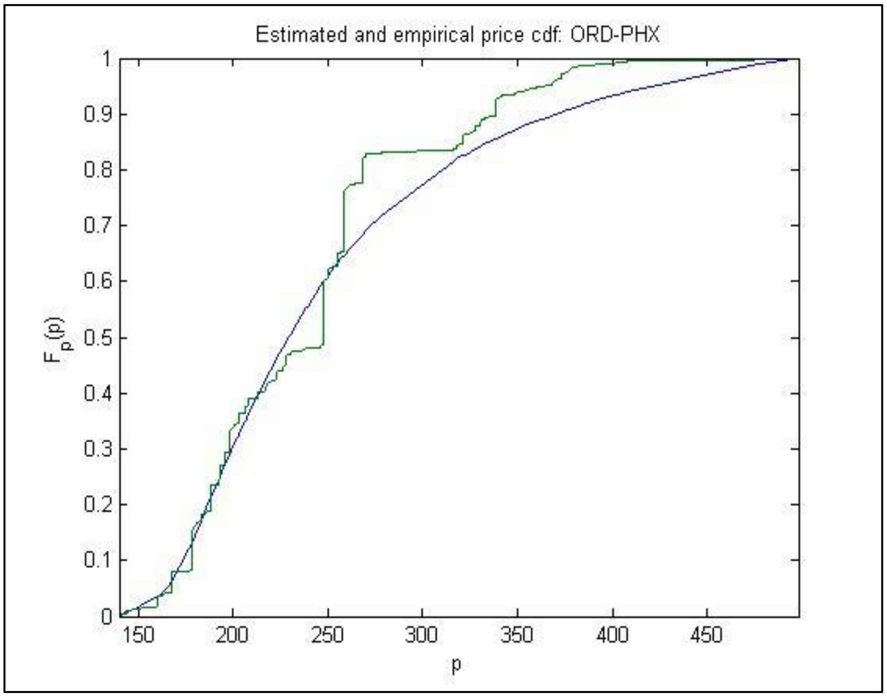

Figure 13 - ORD-SFO Non Sequential Estimation

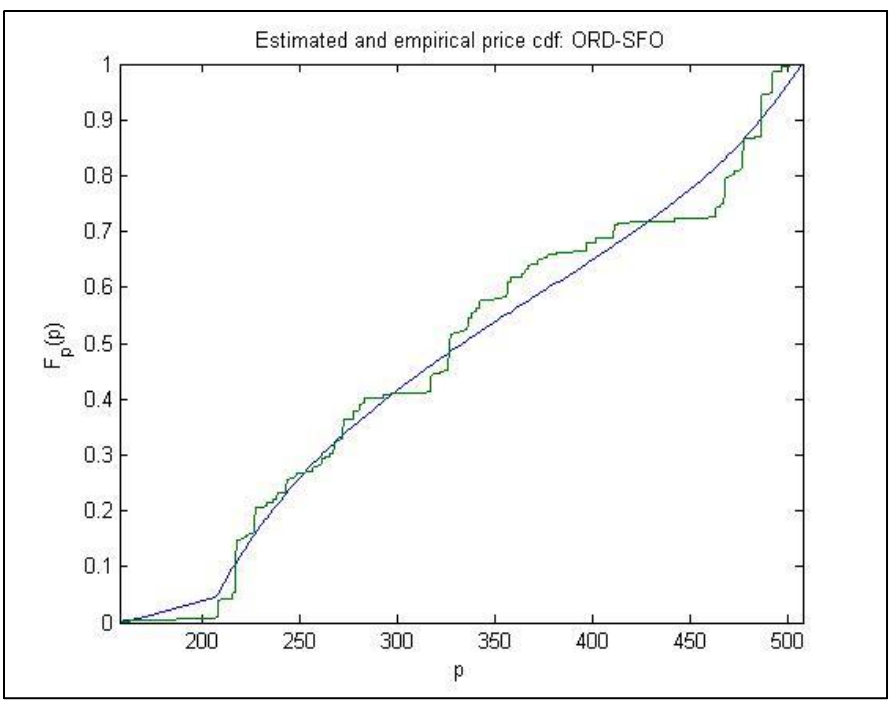

Table 1 - Summary Fare Statistics by Route

Route		Obs	Mean	Standard Deviation	Median	Min	Max
ATL - ORD	All	1,073	186.2	28.0	180.0	133.0	303.0
	Offline	953	188.1	28.5	183.0	133.0	303.0
	Online	120	171.0	44.6	166.0	139.0	234.0
DTW - LGA	All	1,141	193.7	49.1	181.9	126.9	296.7
	Offline	928	189.8	33.8	217.8	126.9	296.7
	Online	213	171.4	37.6	155.9	129.0	280.0
ORD - PHX	All	954	240.6	60.2	248.0	140.2	498.0
	Offline	854	244.0	60.9	248.0	140.2	498.0
	Online	100	211.6	44.6	198.0	142.5	338.0
ORD - SFO	All	894	341.8	101.7	326.5	158.3	507.7
	Offline	790	351.0	101.8	337.1	158.3	507.7
	Online	104	271.8	68.8	266.3	163.0	492.2

Table 2 - Summary Fare Statistics by Time Window

Route	Day window	Obs	Mean	Standard Deviation	Median	Min	Max
ATL - ORD	8 -10	463	189.2	28.6	183.0	133.1	303.0
	11 - 14	610	183.9	27.3	180.0	133.1	303.0
DTW - LGA	8 -10	460	194.0	38.8	185.7	126.9	292.6
	11 - 14	681	193.5	34.2	182.0	129.0	296.7
ORD - PHX	8 -10	370	240.7	59.5	248.0	140.2	473.1
	11 - 14	584	240.5	60.7	248.0	142.5	498.0
ORD - SFO	8 -10	376	351.9	101.1	341.5	162.2	497.3
	11 - 14	518	334.4	101.6	322.2	158.3	507.7

Table 3 - Sequential Model Estimates

Route	θ_1	θ_2	Mean Search Cost	Median Search Cost	α	$F_c^{-1}(1-\alpha;\theta)$
ATL - ORD	0.012	0.293	123.8	3.4	0.1457	112.4
DTW - LGA	0.04	0.331	246.5	13.2	0.2544	103.3
ORD - SFO	0.034	0.256	209.2	8.1	0.2188	174.3
ORD - PHX	0.025	0.305	215.2	7.5	0.1311	255.5

Table 4 - Non-sequential Model Estimates

Panel A - ML Estimates of the Non-Sequential Search Model				
	ATL - ORD	**DTW - LGA**	**ORD - PHX**	**ORD - SFO**
p_1	133	127	140	158
v	303	297	498	508
N	72	90	111	118
M	1,073	1,141	954	894
q_1	0.25 (0.02)	0.21 (0.06)	0.21 (0.03)	0.35 (0.05)
q_2	0.66 (0.01)	0.26 (0.03)	0.50 (0.08)	0.10 (0.02)
q_4	.	.	.	0.45 (0.02)
q_6	.	.	0.13 (0.31)	.
q_7	.	0.34 (0.05)	0.06 (0.33)	.
.	.	.	.	.
q_{72}	0.09 (0.03)	.	.	.
q_{90}		0.19 (0.16)	.	.
q_{111}			0.10 (0.09)	.
q_{118}				0.09 (0.07)
r	127.62 (0.42)	125.19 (0.52)	135.52 (0.81)	148.74 (1.37)
avg mark-up	0.304	0.329	0.404	0.523
LL	4,805.00	5,651.46	5,166.86	5,058.05
KS	3.26	5.52	4.28	2.65

Panel B - ML Estimates of the Non-Sequential Search Model				
Δ_1	18.38 (0.31)	25.05 (2.11)	41.80 (7.16)	58.35 (0.59)
Δ_2	6.43 (0.25)	11.50 (2.15)	15.54 (2.55)	28.08 (1.02)
Δ_3	3.31 (0.17)	6.67 (1.57)	8.24 (2.36)	15.89 (0.93)
Δ_4	2.11 (0.12)	4.36 (1.14)	5.25 (2.35)	10.05 (0.75)
Δ_5	1.53 (0.10)	3.08 (0.85)	3.72 (2.21)	6.89 (0.60)
Δ_6	1.20 (0.08)	2.28 (065)	2.83 (2.03)	5.04 (0.48)
Δ_7	1.00 (0.07)	1.76 (0.51)	2.25 (1.84)	3.87 (0.40)
Δ_8	0.86 (0.06)	1.41 (0.39)	1.86 (1.76)	3.09 (0.33)
Δ_9	0.76 (0.06)	1.15 (0.33)	1.58 (1.60)	2.55 (0.29)
Δ_{10}	0.69 (0.05)	0.96 (0.27)	1.36 (1.46)	2.16 (0.25)
SC mean	8.84	8.85	17.05	27.75

Table 5 - Search Costs Estimates Offline and Online

Route	**SC offline** (US$ - % mean route fare)	**SC online** (US$ - % mean route fare)
ATL - ORD	9.19 (4.9%)	5.05 (2.7%)
DTW - LGA	16.74 (8.6%)	9.10 (4.7%)
ORD - PHX	22.20 (9.2%)	13.00 (5.4%)
ORD - SFO	35.03 (10.2%)	18.26 (5.3%)

APPENDIX B

DERIVATION OF LIKELIHOOD FUNCTION FOR THE SEQUENTIAL MODEL

Given α and r, we can estimate the τth quantile of the reservation price distribution, denoted $G_{\bar{p}}^{-1}(\tau;\alpha,r)$, using the indifference condition (8):

$$(\overline{p}-r)\alpha = (G^{-1}(\tau;\alpha,r)-r)(1-\tau) \Leftrightarrow G^{-1}(\tau;\alpha,r) = \alpha\frac{(\overline{p}-r)}{1-\tau}+r$$

Let $F_c^{-1}(\tau;\theta)$ dente the τth quantile of the parameterized cost distribution, where θ denotes the parameters of this distribution that we wish to estimate. By the consumers' reservation price condition, we know that

$$F_c^{-1}(\tau;\theta) = \int_{\underline{p}}^{G^{-1}(\tau;\alpha,r)} F_p(p)dp$$

And therefore

$$(F_c^{-1})'(\tau;\theta) = F_p(\frac{(\overline{p}-r)}{1-\tau}+r)\frac{(\overline{p}-r)\alpha}{(1-\tau)^2}$$

In what follows, let $c(\tau;\theta)$ denote the τth quantile of $F_c(\tau;\theta)$ $(i.e., F_c(c(\tau;\theta);\theta) = \tau)$. Changing variables from τ to $p \equiv \alpha[(\bar{p}-r/1-\tau)]+r$ we can derive the price CDF corresponding to θ, α, and r,

$$F_p(p;\theta) = \frac{\alpha(\overline{p}-r)}{(p-r)^2 * f_c(c(1-\alpha\frac{\overline{p}-r}{p-r};\theta);\theta)}$$

with a corresponding density function $f_p(p;\theta)$ that can be derived by differentiating the above with respect to p:

$$f_p(p;\theta) = \frac{2\alpha(\overline{p}-r)}{(p-r)^3 * f_c(c(1-\alpha\frac{\overline{p}-r}{p-r};\theta);\theta)} - \frac{\alpha(\overline{p}-r)^2 * f_c^{'}(c(1-\alpha\frac{\overline{p}-r}{p-r};\theta);\theta)}{(p-r)^4 * [f_c(c(1-\alpha\frac{\overline{p}-r}{p-r};\theta);\theta)]^3}$$

The maximum likelihood estimates for the θ parameters are estimated by maximizing the sample log-likelihood function $\sum_i \log f(p_i;\theta)$. The variance-covariance matrix of the estimates is approximated by the inverse of the sample analog of the outer product of the gradient vector:

$$V = [\sum \frac{\partial log f_p(p_i;\theta)}{\partial\theta}\frac{\partial log f_p(p_i;\theta)}{\partial\theta'}]^{-1}$$

where the gradient vector for each observation *i* is, in turn, approximated by numerical derivatives.