A Model of Optimal Consumer Search and Price Discrimination in the Airline Industry

David Liu

Sunday 15$^{\text{th}}$ November, 2015

Abstract

The welfare effects of price discrimination in the market for airfare can be ambiguous: price discrimination can increase airline revenue, but it can also allow for a more efficient allocation of tickets. This paper examines the market for airfare with a structural model of consumer search. My model features sophisticated consumers with rational beliefs about future price movements. Consumers in the market make optimal decisions based on their beliefs about future prices, search costs, and their probability of flying. I model beliefs about future prices as a Markov process based on flight characteristics and current prices. Using data on daily price and quantity in monopoly markets, I estimate the demand for airfare and calculate consumer welfare. I find that leisure travelers account for 70% of all air travel and are more prone to searching for tickets, an average of 2.48 times compared to 1.03 times for business travelers. I compare welfare in a counterfactual market in which airlines cannot price discriminate. The results show that without price discrimination, 5.01% fewer tickets are sold, and consumer welfare increases as a result of lower prices, less search and less cancelation fees.

1 Introduction

The price of airline tickets fluctuates over the entire period of sales, with prices typically rising as the departure date arrives. The price trend is attributed to intertemporal price discrimination: heterogeneous consumers enter the market at different times, creating incentives for airlines to price discriminate. Consumers with higher valuations for flying tend to enter the market near the departure date, so airlines naturally raise prices to coincide with times of high demand. The

welfare effects of price discrimination in this setting can be ambiguous, because price discrimination not only allows for higher prices but also allows for more efficient allocation of tickets. Price discrimination also induces more consumer search and associated search costs, which factor into welfare.

The goal of this paper is to study the welfare effects of intertemporal price discrimination in the market for airline tickets. I model how forward-looking consumers search for airline tickets and estimate the welfare effects of intertemporal price discrimination on the estimated consumer population. My model features consumers who have rational beliefs for how ticket prices may change in the future, which they combine with beliefs on their probability of flying as well as search costs to make a decision on what optimal day to search for tickets. Using price and quantity data on monopoly routes, I estimate demand and welfare under alternatives without price discrimination. I find that price discrimination's effect on allocation efficiency is overshadowed by the higher prices faced by consumers, resulting in a net decrease in consumer welfare.

Two features of this market set it apart from commonly studied durable goods markets. One feature that makes this analysis tractable is that airline tickets have a fixed consumption date. This allows me to use backwards induction to solve for optimal strategies for both airlines and consumers. A second feature is the capacity constraints of airlines. In the short run, airlines have negligible marginal costs, and their objective is to maximize revenue by selling out on tickets for each flight. As a result, idiosyncratic shocks to demand can cause airfare to fluctuate over a short period. These price fluctuations give insight into consumer search behavior and the firm's profit maximization problem. In addition, the limited capacity means the declining price model commonly seen in durable goods models (Board and Skrzypacz (2013)) does not necessarily apply.

For my primary data, I collect daily flight prices over the thirty day period prior to departure and estimate daily quantities sold by scraping the seat maps as well as comparing the number of tickets available in the fare class. I focus on small markets which have routes with one or two flights per day, in order to study the problem in the simplest setting.

A novel aspect of this paper is that I estimate a sophisticated model of consumer behavior through forward-looking consumers who have rational beliefs about price behavior. In my structural model of consumer search, consumers enter the market looking to purchase a single ticket for a route on a specific day. After observing prices, they make the decision between purchasing a ticket for the route immediately, waiting and searching again in a future period, or exiting the market. Consumers base their purchasing decision on how prices will evolve, which I estimate as a Markov process where the state variables are current prices. The movement of prices in the future depends on the price of tickets today, number of flights in the market, number of days remaining,

and date effects. Prices are very likely to increase on certain days before departure, particularly on 21, 14, 7 and 3 days before, and consumers are able to predict this.

Each consumer has individual beliefs about the probability that she will fly, and those beliefs either converge to 1 over time or the individual exits the market. Consumers are likely to search again later under one or more of the following conditions: if they believe prices will be lower in the future, if they will become more certain about flying (and thus less likely to have to cancel a purchased ticket), or if their individual search costs are low.

To estimate my model, I use a maximum simulated likelihood estimator. My model uses parameters on the distribution of consumer characteristics, arrival rate, and flight preferences to match predicted ticket sales with actual sales. Using these parameters, I simulate individual consumers drawn from the parameterized consumer distribution, and for each consumer I calculate their optimal choice in each period after observing any combination of prices. By adding the parameterized arrival rates, I can calculate a distribution of the number of tickets sold under my model in each period conditional on past prices and quantities. I then calculate the parameters that maximize the likelihood of observing the empirical data.

My estimates indicate that two very different populations of consumers enter the market at different rates. A large mass of travelers with high price sensitivity and low search costs enter the market very early: This population will choose to search for the lowest priced tickets. The second population of travelers arrive throughout the life of the flight and very rarely search: they decide to either immediately purchase a ticket or exit the market. I call these two populations leisure and business travelers, respectively.

I find that leisure travelers account for 70.8% of all air travel, and pay $66 less on average for a ticket compared to business travelers. Search is three times more costly for business travelers than leisure travelers, and as a result, leisure travelers spend more time searching: leisure travelers who fly will choose to search an average of 2.48 times, compared to 1.03 times for business travelers.

To estimate the effects of price discrimination, I modify my model to have airlines set a single price for the lifetime of the ticket. Since airlines cannot dynamically adjust prices to fill the plane, I find that the average price of tickets bought decreases from $226 to $210. Since consumers know that prices will not change, they have lower incentives to search and there are gains to consumer welfare in saved search costs, lower prices and few ticket refunds. In another counterfactual, I examine the welfare effects of giving airlines more information: If airlines had full information on the consumers in the market, I find large gains in both consumer welfare and airline profits.

This paper is related to several strands of literature. In airline price discrimination, my model

most closely follows those of Lazarev (2013) and Williams (2013).

Lazarev estimates a model of consumer demand featuring consumers who are uncertain about flying. His consumers maximize their expected welfare from flying by calculating their optimal period to purchase tickets, which depends on deterministic future prices as well as the consumer's own probability of flying. To estimate his model without daily quantity data, Lazarev uses a sophisticated series of moment conditions combining daily price data and quarterly aggregated quantity data. I expand on his model by putting uncertainty into the market for airfare. Airlines are uncertain about future demand, and as a result consumers are unable to perfectly predict future prices. I am able to develop in this way by incorporating day-level quantity data. The finer level of data allows me to model the individual consumer's search behavior and gives consumers rational expectations and uncertainty about future prices.

Williams (2013) estimates demand for airfare and the airlines' optimal pricing policy using day-level price and quantity data. In his model, consumers decide between purchasing a ticket immediately or exiting the market. He calculates optimal pricing strategies for airlines which depending on time and number of tickets remaining. He finds that in addition to price discrimination, dynamic price adjustment is also important for airline revenue maximization. One of his conclusions is that consumers who fly with certainty will never delay purchasing. My model follows the approach Williams but features consumers who are initially uncertain about their probability of flying and hold rational beliefs about future prices. Consumers may search for airline tickets over time to find the optimal period to purchase tickets. Another paper, Etzioni et al. (2003) finds that there are substantial potential savings for sophisticated consumers who search for tickets.

Several other papers study the price dispersion in airlines (Borenstein and Rose (2007), Gerardi and Shapiro (2009), Clemons et al. (2002)). My paper also relates to studies of why firms use intertemporal price discrimination. Varian (1980) presents a search based model to explain price fluctuations as mixed strategy equilibria. Hendel and Nevo (2011) find price discrimination in storable goods market to be caused by the existence of both naive and sophisticated consumers.

The paper proceeds as follows. In the next subsection I present motivating facts about the market for airfare. Section 2 presents the model of consumer choice. Section 3 describes the data used in the model and presents the Markov model for how consumers predict future prices. Section 4 describes the maximum likelihood estimation approach and estimation process. Section 5 presents the estimation results. Finally, section 6 presents two counterfactual alternatives to price discrimination.

2 Model

In this section, I first describe the factors that affect consumer preferences. I then present my model of how consumers make their optimal purchasing decision.

2.1 Consumer Preferences

I model a world with two populations of consumers, indexed by $i = 1, 2$[1]. Every period, some consumers from each population enter each market, with the number of entrants drawn from a poisson distribution with parameter λ_{it}. Consumers in a market are only interested in purchasing a single ticket in a specific market, which I define as a directional city pair and date of departure. Each consumer who enters the market will observe prices of flights for that day and can choose to either purchase a ticket on any flight in the market or come back in any future period to observe prices again.

Consumers are initially uncertain about whether they will fly or if they will ultimately cancel their plans. As time passes, the uncertainty decreases. Consumer j who enters the market in period t believes that she will fly with probability x_{jt}; On each future date, there is a known probability $0 \leq \rho_{jt} < 1$ of her realizing that she will not fly, and x_{jt} represents the probability that she never drops out:

$$x_{jt} = \Pi_{\tau=t+1}^{T}(1 - \rho_{j,\tau})$$

In the next period, $x_{j,t+1} = x_{j,t}/(1 - \rho_{j,t+1})$ if consumer j does not drop out, so $x_{j,t}$ is weakly increasing over time conditional on j being in the market. As the date of the flight approaches, each consumer will either exit or their x_{jt} grows to one. I parameterize the path of $x_{jt}, ..., x_{jT}$ in a way that is simple and flexible using three parameters for each population. A consumer of population i has initial chance of flying $x_{j0} \sim N(\mu_{x,i}, \sigma^2_{x,i})$. I use parameter η_i to describe how quickly $x_{j,t}$ converges to 1. Given η and $x_{j,0}$, $x_{j,t} = (t/30)^{\eta}(1 - x_0) + x_0$ or alternatively, $\rho_t = 1 - \dfrac{(t/30)^{\eta}(1 - x_0) + x_0}{(t + 1/30)^{\eta}(1 - x_0) + x_0}$.

If a consumer decides to fly, they get utility based on their preferences over flight characteristics and the price of the ticket. Consumer j of population i who buys a ticket on flight k in period t at price p_{kt} and still wants to fly on the final date has the following utility from flying:

$$\mu_{jkt} = \alpha_j + \beta_i p_{kt} + \gamma_j X_k + \varepsilon_{jk}$$

[1] My two type consumer model follows past studies in the airline industry(Berry and Jia (2008), Lazarev (2013)). In the appendix I consider alternatives to this assumption

X_k includes flight characteristics: day of week, flight time, carrier fixed effects, route fixed effects. ε_{jk} are distributed i.i.d. type 1 extreme value which represent consumer's idiosyncratic preferences and do not vary over time. Each population has a preferred time and day of week to fly, and their utility is based on $\gamma_j^{time}|\text{Flight Time} - \text{Optimal Time}_i|$. Within a population i, preferred times to fly are drawn from a normal distribution with estimated mean and variances. Carrier and route fixed effects are common to all consumers and between populations. β_i is the consumer's population's price sensitivity and γ_j are the individual idiosyncratic preferences over flight characteristics and are normally distributed with parameters for mean and variance. Consumers within a population have a common β_i but individually varying x_{jt} and preferences over flight characteristics.

In period t, consumer j believes that she will fly with probability x_{jt}, so her expected utility from purchasing a ticket in that period is the weighted utility from flying and not flying. If a consumer purchases a ticket and then chooses not to fly, they have the options of changing or canceling the ticket. Major carriers almost universally charge \$200 to change a ticket, while some smaller carriers like Alaska Airlines charge less. More expensive tickets can be canceled, but tickets that cost less than \$200 become worthless if the consumer realizes they will not fly. Total utility from buying a ticket at time t becomes

$$u_{jkt} = x_{jt}\mu_{jkt} + (1 - x_{jt})(\beta_i max(p_{jt} - Fee_k, 0))$$

2.2 Consumer's Decision

Given the consumer's utility function, I can calculate the consumer's optimal decision in any period. Upon entering the market for the first time, the consumer immediately observes the prices of all tickets in the market. After observing prices, the consumer has three choices: Buy a ticket immediately, return in a future period τ and search again, or exit the market. If a consumer chooses to search again, they pay search cost c which represents time and energy spent searching. Individual cost from waiting is distributed from a normal distribution truncated below by 0. The forward-looking consumers base their decision on their beliefs about future prices, the path of x_{jt}, and their search cost. By choosing to return in period τ, there is a $1 - (1 - \rho_{j,t+1})...(1 - \rho_{j,\tau})$ probability that at $t = \tau$ the consumer will realize they will not fly and will not pay c.

By choosing to search in a future period, the consumer knows that she will become more certain about flying but incurs cost c as well as potentially facing higher prices or even a sold out flight. Intuitively, returning in the future to search is preferable if consumers believe prices will not increase, their x_{jt} will increase, and c is small.

Let V_{jt} be the consumers valuation for being in the market in period t conditional on $x_{jt} \neq 0$. V is the utility payoff of the consumer's optimal action at period t, and we can write it as the maximum expected payoff among all of the consumer's choices (buy a ticket, come back in any future period, or exit the market).

$$V_{jt} = \max_{k,t'} \left(u_{jkt}, \frac{x_{jt}}{x_{jt'}} \left(E[V_{jt'}] - c \right), 0 \right)$$

$E[V_{j,t+n}]$ represents the consumer's expected valuation for being in the market in a future period given what they know in the present conditional on having $x_{j,t+n} \neq 0$. Consumers calculate this based on their beliefs of how prices will evolve from observed prices today, as well as the path of x_{jt}.

Consumer j purchases a ticket on flight k immediately in period t if it gives a higher utility than buying on another flight k' and higher utility than waiting and searching again in any future period. This is represented by the inequalities:

$$\begin{aligned} u_{jkt} &\geq 0 \\ u_{jkt} &\geq u_{jk't} \text{ for all } k' \\ u_{jkt} &\geq \frac{x_{jt}}{x_{jt'}} (E[V_{jt'}] - c) \text{ for all } t' \end{aligned}$$

If $\frac{x_{jt}}{x_{jt'}}(E[V_{jt'}] - c) > u_{jkt}$ for any t', the consumer believes that the expected payoff from searching again in period t' is greater than purchasing a ticket immediately, so they will choose to return in some future period and search again: the consumer picks the period $\hat{t}$ where their $\frac{x_{jt}}{x_{j\hat{t}}}(E[V_{j\hat{t}}] - c)$ is largest: if $\frac{x_{jt}}{x_{j\hat{t}}}(E[V_{j\hat{t}}] - c) \geq \frac{x_{jt}}{x_{jt'}}(E[V_{jt'}] - c)$ for all t' then return to search again in period $\hat{t}$. This model allows consumers to return and search repeatedly if their observed prices lead them to believe future prices will be low, or consumers may buy on their first search if their search costs are high or expected future prices are high.

This model of consumer search allows me to calculate an individual's decision to purchase at each point in time. Combined with the population of each consumer type, and the distribution of consumer characteristics, I can calculate Q_{kt}, the expected number of tickets sold on flight k in period t.

3 Data

In this section, I describe my data sources, and present summary statistics and the results of my estimation of price path movement.

3.1 Data Sources

My data selection criteria require routes that are served by a single carrier and are primarily flown as direct routes rather than as a leg of a longer flight. To identify these markets, I use the Bureau of Transportation Statistics DB1B database.

The DB1B is a 10% sample of airline tickets from US operating carriers. It includes data on ticket-specific information such as number of connecting flights and number of passengers on each ticket. By looking at ticket sales in the first two quarters of 2013, I identify all routes that satisfy the following criteria: I look for routes where over 70% of travels starts start from the origin and end their trip at the destination (complete flights instead of legs of a longer trip) and 70% of all travel from the origin to the destination is through direct flights. These route selection criteria rule out two potential problems: by focusing on routes that are complete flights, I eliminiate the difficulty of calculating how airlines and consumers value legs of flights; and by using routes where the majority of travelers fly directly, I minimize the effect of competition with other carriers offering indirect routes. In addition, I look for routes that are monopoly, with one or two flights per day: all markets in my data have a single carrier that operates one or two flights on the day. Examples of monopoly routes that satisfy my criteria are MIA (Miami, FL) to IAD (Washington, DC) operated once a day by American Airlines, or JAX (Jacksonville, FL) to JFK (New York, NY) operated twice a day by Jet Blue.

I identify 55 routes that satisfy my criteria, 23 of which have days with single flights and 46 with days with two flights. A list of all routes in my data appear in appendix **??**. On those routes, I find a total of 319 day$\times$route pairs which have one or two flights, 193 with two flights and 126 with one, for a total of 512 flights spanning a two week period of departure in February 2015. I define each of these day$\times$route pairs as a market, and for each market I collected price and quantity data for 30 days leading up to the departure. I have a total of 12804 price and seatmap observations. Of my 512 flights, I find that 308 sold out before the final date of departure, and past that date no data was observable.

To get flight cost and quantity data, I use the internet travel website Expedia. I collect daily data on prices for all routes that fit my criteria. Expedia displays a seat map for some flights

indicating which seats are available. Using changes in the seat map for each flight, I get a proxy for the number of tickets sold each day. In addition, I gather a second measure of quantity through each flight's basis code. Expedia reports the number of available seats remaining for the lowest cost basis code available for sale, and daily changes to that number suggest a number of ticket sales. The two sources combined give a measure of number of tickets sold each day: If either a seat fills in the seat map or the number of available tickets in a fare class decreases, I can infer a ticket sale. A potential problem of this measure of quantity is the consumers who purchase tickets but do not select a seat immediately, but instead pick it at the time of flight or check in. These consumers will appear in the seat map as a last minute purchase one or two days before the flight, but they will decrement the quantity of tickets available in their fare class. This may bias my results toward finding more consumers searching and purchasing in the final days before departure.

In addition, Expedia on some dates did not provide seat maps on certain flights, leave short gaps in my data. 114 flights have full data through the entire 30 days, and the remaining 398 are missing data points either through selling out or missing seat maps. From these 12,804 observations, I had 11,592 pairs of days of consecutive days in which I could calculate the number of tickets sold.

3.2 Summary Statistics

The primary feature of interest of the airline ticket market is intertemporal price discrimination. Figure 1 shows the movement of the average price of airfare over the 30 days before departure. Prices generally trend up, with predictable, discrete jumps in average price after 7 and 14 days before departure, and smaller increases around the 21st day before departure. Of particular interest is that average price jumps by \$10.47 after the 21st day before departure, \$37.4 after the 14th day and \$37.37 after the 7th day.

Figure 2 shows the average number of sales observed on each day before flight departure. I observe the large numbers of purchases 7 and 14 days before departure, which correspond to the final days before a price increase. This is suggestive of consumers choosing to purchase before a predicted price increase, and my search model will attempt to explain this behavior.

Individual flight prices do not perfectly follow the trend: prices fluctuate even between flights by the same carrier, and price increases can be unpredictable. Within my 512 flight data set, I found 1121 instances of price decreases, and 1482 price increases. Of the 512 flights, 308 withdrew from Expedia prior to the final day of departure, indicative of the flight selling out. Figure 3 shows the price movement of prices in four markets. Flight prices in markets with two flights usually move in parallel: when both flights still have tickets available for sale, the price action for the two

Figure 1: Average Prices of tickets over time

Figure 2: Average Number of tickets sold over time

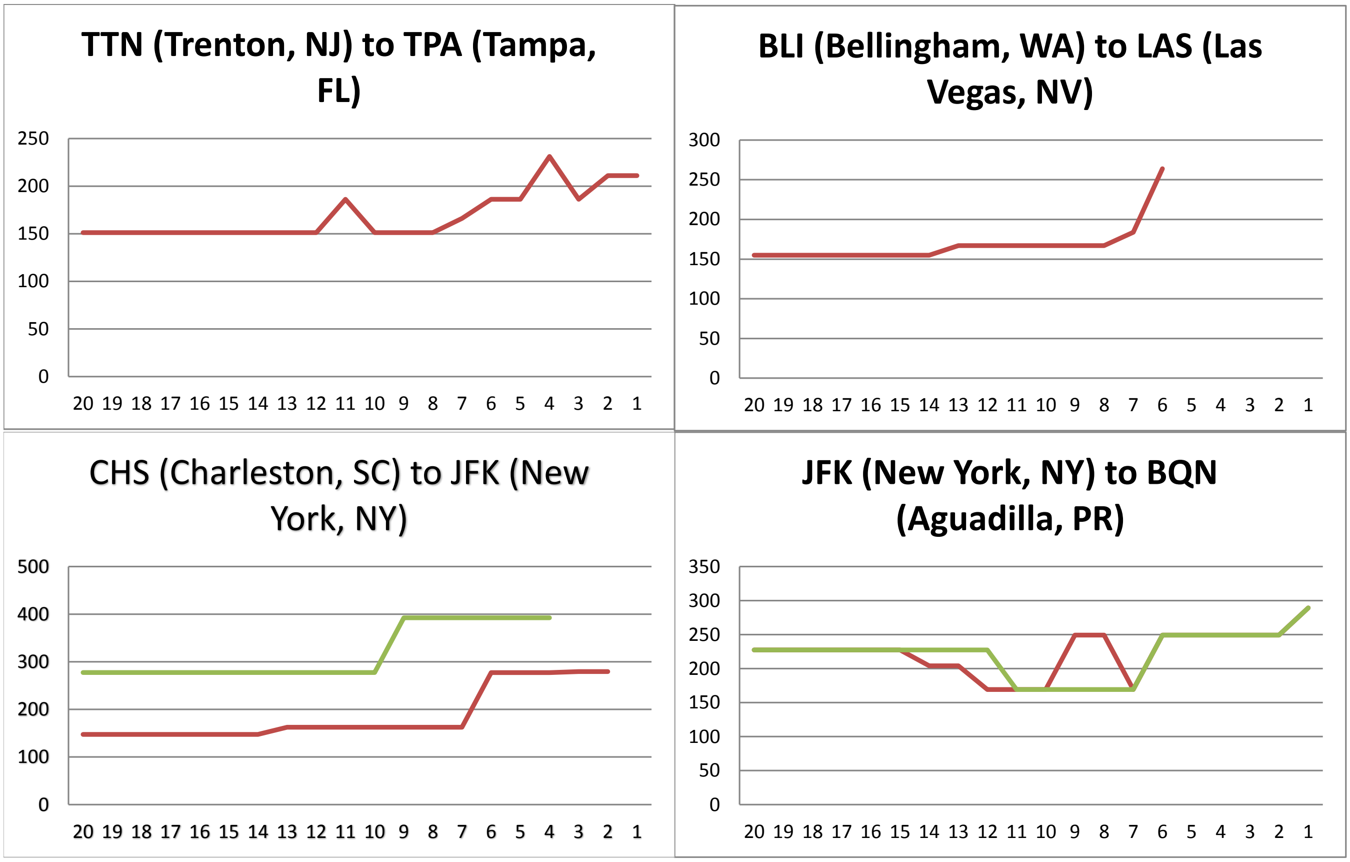

Figure 3: Price Movement for four flights

flights in each period have a 82.3% chance of matching, where the actions are increase, decrease, stay constant, or sell out.

Because I am restricting to monopoly routes with up to two flights per day, my routes are shorter than average flights in markets with fewer consumers, and is not representative of airline sales in general. The average price of flights in my data is $232.5, with flights seating an average of 57.7.

3.3 Markov Price Path

I model consumers as having rational expectations of future price movement. I estimate their beliefs as a 10 state Markov process with transition probabilities that vary over time and across markets. Transition probabilities depend on the price of tickets today, number of days until departure, and current day of week, and dummy variables for 3, 7, 14, and 21 days remaining. Based on these parameters, In each period I estimate the probability that prices will move up, down, remain the

same, or the flight will sell out. Conditional on a predicted price change, the amount of change follows the empirical distribution of price changes for that price bin derived from all flights in my data set. In my model of consumer beliefs, future prices do not depend explicity on the number of seats remaining on a flight. I test this assumption in the appendix and find that after conditioning on all of the independent variables listed above, the number of seats remaining on a flight does not help predict the number of seat sales, which suggests that the number of days remaining and current prices are sufficient to model consumer beliefs.

In markets with two flights, prices can move either in parallel with probability q or independently of the price of other tickets in the market with probability 1-q (In the case of independent movement, prices may still change in exactly the same way by chance). To make the computation process feasible, I divide prices into 10 bins such that each flight's highest and lowest bins are the highest and lowest observed price on that route, and remaining bins are divided evenly between the two extremes. I predict price movement between bins.

The probability of each i (increase, decrease, unchanged, or sell out) follows a logit model, and I estimate separate parameters for the probability of price increase, decrease and selling out conditional on independent movement and correlated movement. In markets with a single flight, price movement uses the estimated coefficients of independent movement. The probability of movement i for flight k is

$$P(i,k) = \frac{M_i}{\sum_k M_k}$$

where $M_{i,k} = \exp(\xi_{i,independent}[p_k, t_k, w, D_k]')$, ξ_i are the coefficients on price of the ticket today, number of days until departure, day of week dummies, and 3, 7, 14, and 21 days remaining dummy variables. I normalize $M_{unchanged} = 1$. I estimate a ξ_i for the case of parallel and independent movement.

In a market with two flights k, k', with probability q both flights will behave in parallel, and with probability $1-q$ flight k will move independently, with the total probability of action i in (increase, decrease, sell out) equal to the

$$P(i,k,k') = (1-q)\frac{M_i}{\sum_k M_k} + q\frac{N_i}{\sum_k N_k}$$

where $N_{i,k,k'} = \exp(\xi_{i,correlated}[p_k, t, w_k, D_k]' + \xi_{i,correlated}[p_{k'}, t_{k'}, w, D_{k'}]')$

Table 1 shows the results of the markov estimation. I find that probability of a parallel movement is 0.72. The coefficient on number of days remaining is positive for selling out, and negative for price gains and losses. As the departure date gets closer, there is an increased chance to sell

out. The coefficients for price gains on the 7 and 14 day dummy variables is large, reflecting the jumps in average prices after those dates. The largest effects of day of week are price increases between Tuesday and Wednesday, and price decrease between Friday and Saturday.

Table 1: Estimated Parameters of Markov Process for Predicted Price Movement

		Independent Movement			Correlated Movement		
	Parameter Estimate	Gain	Loss	Sell Out	Gain	Loss	Sell Out
	Constant	0.015	-4.741	0.045	-1.600	-8.639	-0.433
	Days Remaining	-0.197	0.000	-0.648	-0.157	0.009	-1.002
	Price(0-1)	0.019	5.130	4.609	0.460	-6.975	0.161
	$Price^2$	0.000	-6.157	-6.657	-2.714	2.863	-1.153
	$Price^3$	-1.662	3.475	3.278	0.942	8.595	2.087
	Price*Days Remaining	-0.017	-0.024	0.007	-0.054	-0.076	-0.102
Days Remaining:	3	0.615	-0.590	-0.283	1.141	0.250	-1.117
	7	2.599	0.346	1.317	3.115	1.091	3.126
	14	3.538	-0.180	0.000	3.718	2.445	0.022
	21	1.885	0.004	-19.169	1.412	-0.038	0.347
Day of Week:	Mon	-0.022	0.119	-1.096	0.278	0.342	0.298
	Tues	1.453	-0.758	1.369	1.681	0.619	1.767
	Wed	-0.175	0.602	0.608	0.232	0.179	0.926
	Thurs	-0.311	0.104	1.942	-0.001	0.475	2.414
	Fri	-0.306	1.127	0.035	0.156	1.892	0.529
	Sat	-0.339	-0.892	0.017	-0.232	0.000	-0.098
Carrier:	AA	-0.212	0.000	0.484	0.493	1.104	2.192
	DL	0.030	0.559	0.287	-0.223	2.760	1.559
	AS	0.054	0.322	0.372	0.277	2.457	1.452
	US	-0.830	0.231	0.236	0.805	1.807	2.930
	B6	-0.006	2.060	0.478	0.680	3.940	2.309
	HA	-0.232	0.682	-0.012	0.647	2.471	2.301

Figures 4 and 5 present a sample of the predicted price movement for two pairs of flights using the estimated Markov movement. In Figure 4, after observing on 2/5 that prices were $200 and $130 for the flights from ANC (Anchorage, AK) to OTZ (Kotzebue, AK), consumers forecast the distribution of possible prices on 2/15 to consider whether to search on that date. There is a high probability that both prices increase, with some possibility of either flight selling out.

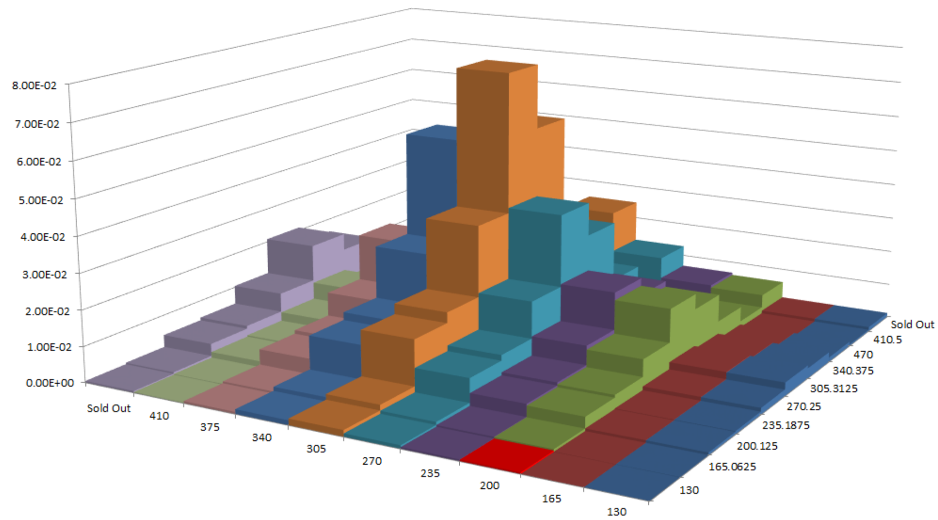

Figure 4: Predicted distribution of prices on 2/15, conditioned on the knowledge that on 2/5 prices are \$200 and \$130 for flights from ANC (Anchorage, AK) to OTZ (Kotzebue, AK).

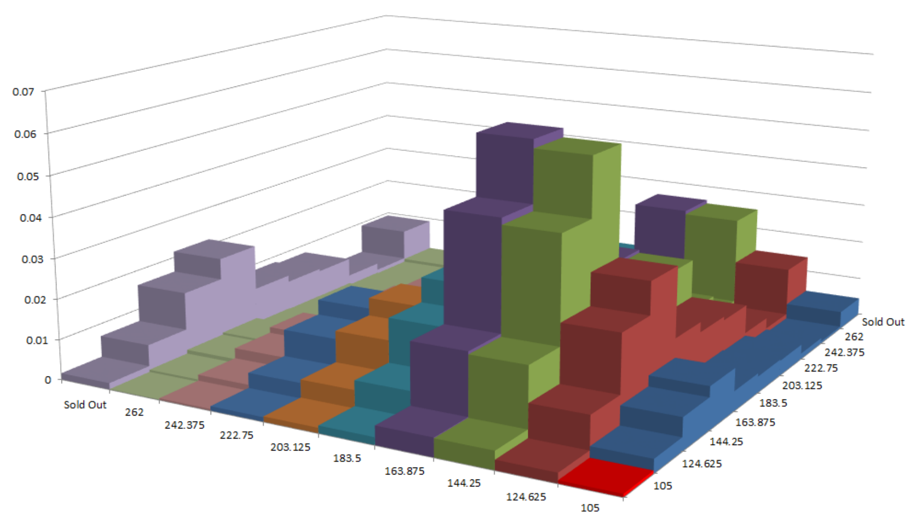

Figure 5: Predicted distribution of prices on 2/15, conditioned on the knowledge that on 2/5 prices are \$105 and \$105 for flights from MCO (Orlando, FL) to CAK (Akron, OH).

4 Estimation

My estimation uses maximum simulated likelihood: for a set of parameters I can calculate the likelihood of drawing the empirically observed quantity of tickets sold in each period. To do this in a computationally feasible way, I simulate draws of potential entrants in each period from the parameterized distribution of consumer characteristics and calculate each simulated consumer's optimal behavior using backwards induction. Then for each period I draw groups of entrants with varying numbers and characteristics (condition on past prices and quantities) to calculate the likelihood of matching the empirical data in each period. In this section I describe in more detail the simulation process.

4.1 Individual Consumers

Because the market for tickets has a final period, I can use backwards induction to calculate each consumer's optimal behavior. I restrict my markets to having a maximum of 2 flights, so each consumer faces at most 2 prices in each period. In the final period $t = T$, consumer's decision to buy is based only on final period prices, and consumer j's condition for buying a ticket on flight k is:

$$u_{jkT} \geq 0$$

$$u_{jkT} \geq u_{jk`T}$$

where u_{jkT} depends on consumer $j's$ individual characteristics. For every pair of prices $p_{kT}, p_{k'T}$, denoted P_T, I calculate consumer's choice and valuation $V_{jT}(p_T)$. Then in the previous period, the consumer observes price vector p_{T-1} and calculates the value for waiting until period T as the expectation $E[V_{jT}(p_T)|P_{T-1}]$, which is weighted by the probability of observing prices p_T given p_{T-1}:

$$E[V_{jT}|p_{T-1}] = \sum_{p_T} V_{jT}(p_T) Prob(p_T|p_{T-1})$$

and so I can calculate the previous period's valuation:

$$V_{j,T-1}(p_{T-1}) = max(u_{jk,T-1}, u_{jk,T-1}, E[V_{i,T}|p_{T-1}] - c)$$

for all possible prices at p_{T-1}. By continuing in this way with backwards induction, I calculate $V_{i,t}(p_t)$ for the consumer in each period for all possible prices.

4.2 Likelihood Estimation

In period t, the number of consumers from population i that enter the market follows a poisson distribution with parameter $\lambda_{it} = f_i(t)$. I estimate each population's $f_i(t)$ as a fourth degree polynomial function.

For computational feasibility, I calculate likelihood in two steps. First, for a given distribution of consumer characteristics, draw consumers from the distribution and I calculate each consumer's optimal purchasing decision in each period after observing the prices seen in the data. Then in each period starting from $t = 1$, I simulate the entry of consumers and calculate the likelihood $L_{m,t}$ that the simulated consumers purchase the same quantity of tickets as the observed data. Each subsequent period features arrivals in addition to the consumers returning from previous periods, where returning consumers come from previous periods' draws that matched the empirical data.

The objective is to maximize the overall likelihood $\prod_m \prod_t L_{m,t}$ across all periods and markets. I search across 95 parameters, including 55 route fixed effects and 6 carrier fixed effects.

To improve computational speed, I run the estimation in two steps. In the outer loop I iterate over the distribution of consumer characteristics. For each distribution of consumer characteristics, I draw 2000 individual consumers from each population and calculate their purchasing decisions conditional on prices in each period. Then in the inner loop I iterate over the arrival rate of consumers to maximize total likelihood conditional on the distribution of consumer characteristics. Consumers entering the market in each period are drawn from the pool of 2000 consumers.

5 Maximum Likelihood Estimation Results

I estimate two populations of consumers and find that they have very different parameters. One population has a high average valuation of flying, high search costs, and low price sensitivity, and the other has a low valuation for flying, low search costs, and high price sensitivity. I call the populations business and leisure travelers respectively. Table 2 lists the estimation coefficients. As a result of their population parameters, I find that the two populations have very different search behavior. Conditional on purchasing a ticket, the average leisure traveller will observe prices 2.48 times, while a business traveler will observe prices 1.03 times. This result comes from primarily from three factors: arrival times, probability of flying, the cost of searching.

One large difference in the two populations is their arrival rates. Figure 6 shows the average arrival rate for business and leisure travelers over the lifetime of the flight. The majority of leisure

travelers start searching for a ticket very early- they observe prices when they initially arrive, and then return to search again. In contrast, business travelers arrive steadily over the lifetime of the flight with a significant rise in the final periods of the flight. Because leisure travelers arrive very early when they may be initially very uncertain about their probability of flying, search is a very important tool for them to maximize their welfare. Business travelers who arrive later in the lifetime of the flight may find that their probability of flying is already high and are more likely to purchase immediately.

The populations have different different probabilities of flying (x_t). On average, leisure travelers have a low initial probability of flying $x_0 = 0.539$, but they receive more information of whether they will fly in each period. Business travelers on the other hand have a high initial probability of flying $x_0 = 0.878$ but receive very little new information each period until very close to the departure date. Figure 7 shows the beliefs about flying for a typical leisure and business traveler.

Business travelers have a high cost to searching, \$73.16, while leisure travelers only experience costs of \$5.27. Since search is very costly for the business travelers, most often they will choose to either purchase a ticket immediately or exit the market and not search again.

These differences in the properties of the two population are the primary drivers of consumer search behavior. Since business travelers do not update their probabilities of flying until late in the life a flight, the only reason that a business traveler would search again would be to wait until the very final periods to purchase a ticket. While flight prices can go down, it is infrequent enough that business travelers never find it worthwhile to search for tickets over time. Leisure travelers' beliefs about their probability of flying increases in each period, so they find it much more worthwhile to delay purchasing a ticket and paying the search costs later. All leisure travelers arrive very early in the process, when their probability of flying is very low. I find that leisure travelers almost never purchase a ticket when first arriving. Figure 8 shows the distribution of consumers of each type in the market for airfare in each period. These include consumers who just entered the market as well as consumers returning from previous periods to search again. Leisure travelers completely exit the market after the 7th day before departure, leaving only business travelers.

As a result of different search behavior, the purchasing patterns of the two populations are very different as well. Figure 9 shows the number of tickets sold to each population as predicted by the model compared to the observed data (conditional on the the flight not selling out). Leisure travelers overwhelmingly purchase tickets on the dates when predicted prices will increase. By waiting until those dates, leisure travelers are most certain about their probability of flying while avoiding the potential price increase on the next day. On the routes in my data where consumers are paying \$220 on average for flights, a \$200 fee to change flights is a large part of consumer welfare

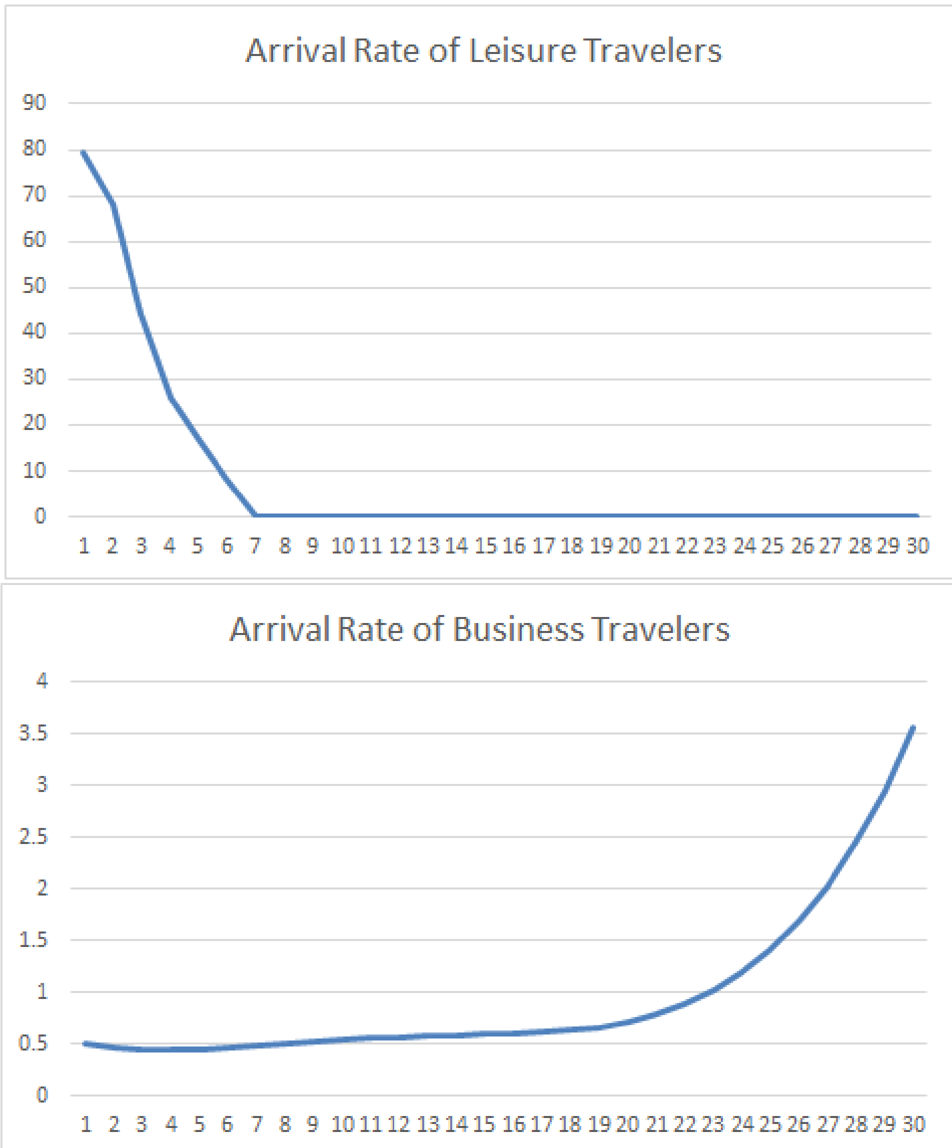

Figure 6: Arrival Rates of Leisure and Business Travelers

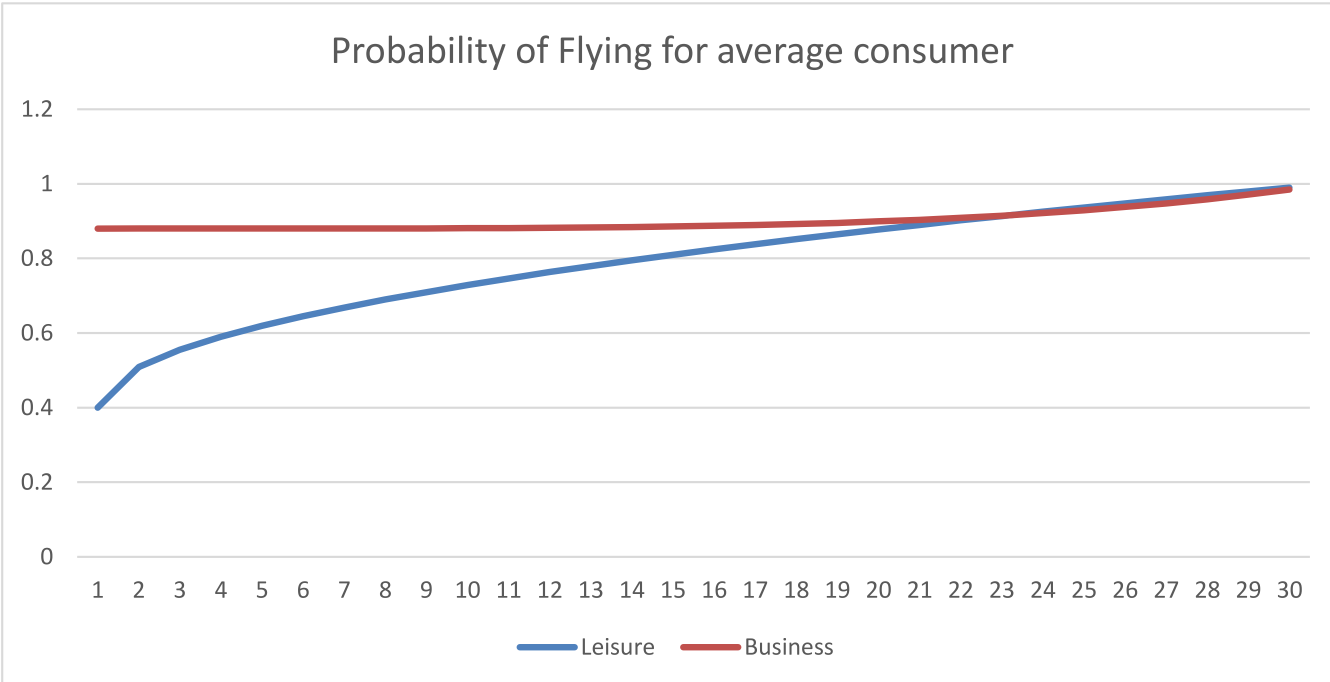

Figure 7: Probability to fly for an average consumer

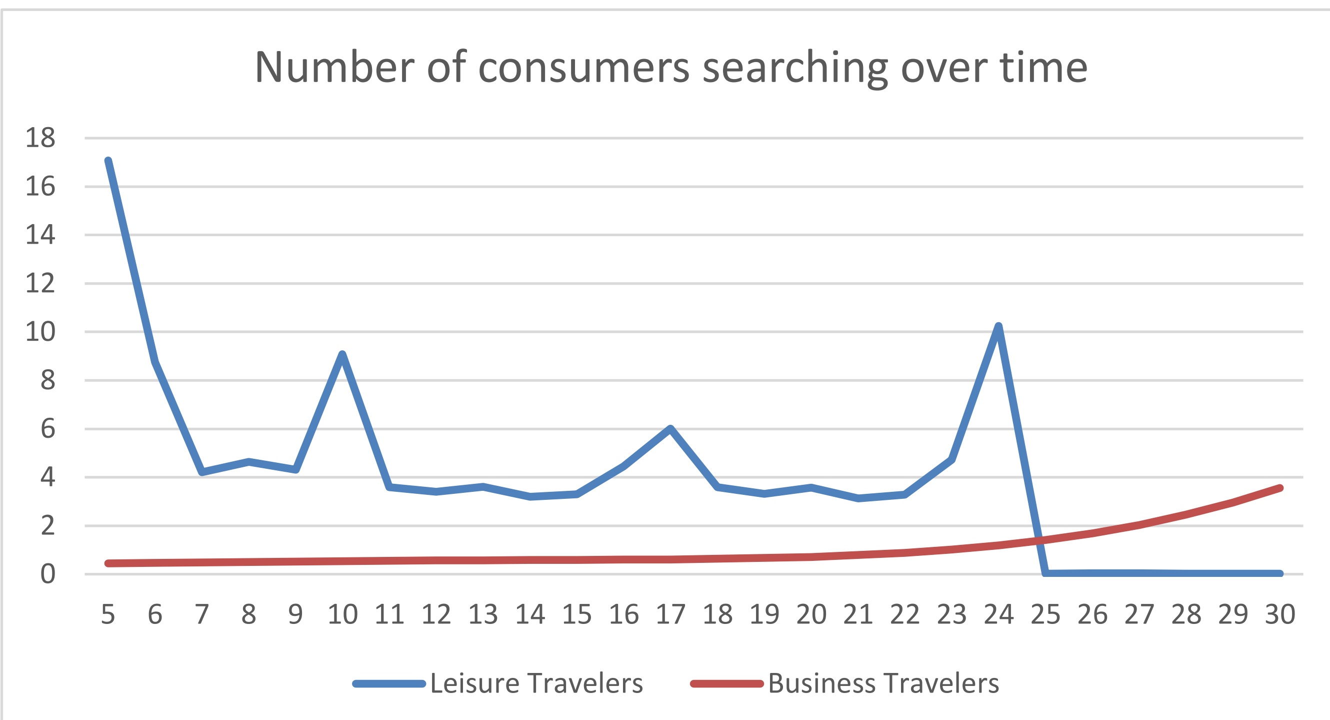

Figure 8: Number of consumers searching over time

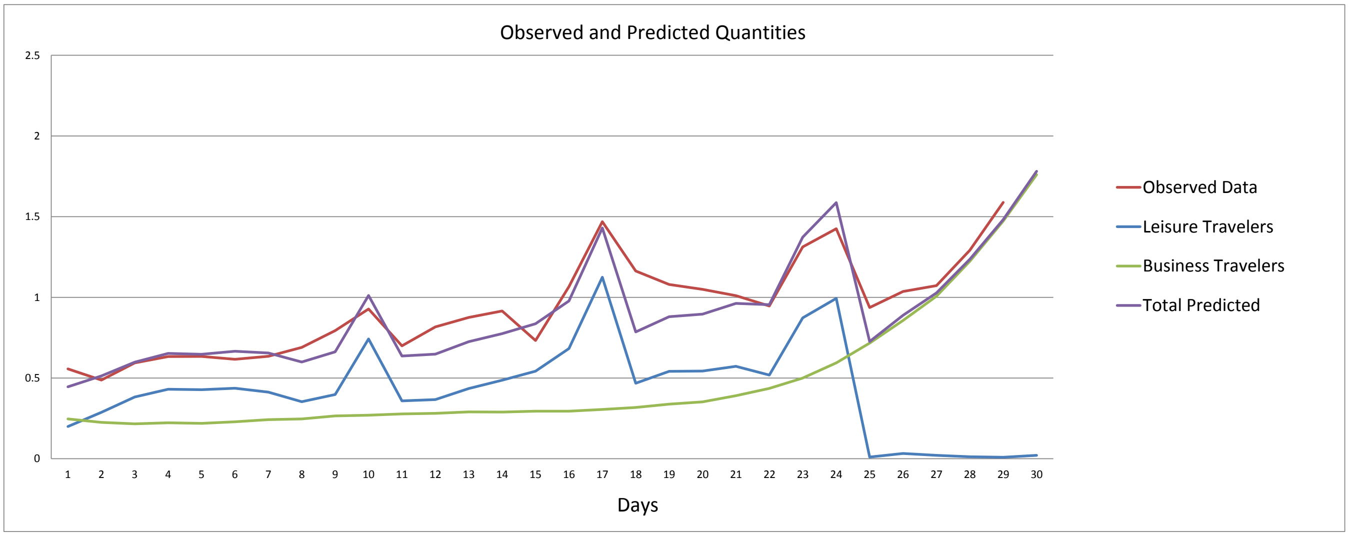

Figure 9: Predicted and Observed Quantity Purchased Over Time

especially for leisure travelers. Because they have low search costs, they would prefer to wait until the final periods to observe prices and purchase a ticket then, but because prices are increasing over time, they feel a pressure to purchase early. Business travelers make purchases throughout the lifetime of the flight, with each passenger purchasing as they arrive. Because leisure consumers are more prone to search and arrive earlier than business travelers, leisure travelers pay lower prices for tickets. The average price for tickets sold was $227.5, and leisure travelers paid $208.22 while business travelers paid $274.31.

While leisure travelers greatly outnumber business travelers, they ultimately only purchase 70.8% of tickets, while business travelers purchase 29.2%. Leisure travelers that enter the market initially often drop out of the market by realizing that they will not fly or having insufficient valuation to purchase a ticket. Because leisure travelers arrive early, it is possible that leisure travelers purchase all of the available tickets and the late arrival business travelers are unable to find tickets. For example the final day before departure on the a flight from JFK (New York, NY) to LGB (Long Beach, CA), at price 280 (higher than the average observed price of $234), if all consumers waited until the final period the purchase, there would be an average of 18.7 business travelers and 14.1 leisure travelers willing to purchase a ticket.

5.1 Airline price estimation

Because my model uses only demand data, I test the supply side of my model by calculating optimal prices set by airlines. Due to the complicated nature of optimal revenue management

Table 2: Maximum Likelihood Estimation Results

	Consumer Parameter	Leisure Travelers Estimate	SE	Business Travelers Estimate	SE
	Constant	243.210	6.998	438.324	24.292
	Price Coefficient	1.003	0.020	0.642	0.078
	Search costs	5.278	0.225	73.163	5.527
	x_0	0.539	0.040	0.878	0.016
	η	0.417	0.052	3.887	0.397
	Preferred Day of Week	6.270	1.258	2.298	0.507
	Day of week coeff	0.016	0.011	0.124	0.056
	Time of Day	9:58	2.424	7:25	2.012
	Time of day coeff	0.014	0.007	0.001	0.001
Individual Variation	Constant	37.397	2.065	164.478	11.803
	Search costs	2.045	0.353	18.133	1.909
	x0	0.165	0.055	0.223	0.040

calculations, I instead calculate airlines' optimal prices in each market if they set a single price for periods of 30-14, 14-7 and 7-1 days until departure. In order to best match how airlines actually set prices, I let consumers beliefs about price movements match my previously estimated maximum likelihood beliefs. I find that the average prices for periods 30-14 and 14-7 match observed prices quite well: in 49 of 55 routes, my predicted optimal prices by the airlines lie within a standard deviation of the mean observed price for the period. But my calculated optimal prices are lower than observed prices for the final 7 days before departure for 20 of 55 routes, which I believe is primarily due to the airline's uncertainty about demand in the final periods. As Williams (2014) observed, an airline's ability to dynamically adjust prices to sell to capacity is a large part of their revenue. In my restricted model where airlines can not changes prices after $t = 7$, I find that airlines set low prices in the final period to ensure sales to capacity. Figure 10 is an example for flights from JNU (Juneau, AK) to KTN (Ketchikan, AK) by Alaska Airlines, where my predicted prices lie within the observed prices except in the final periods.

6 Counterfactual Simulations

In this section I consider several alternate models to evaluate the welfare effects related to intertemporal price discrimination. First, I force airlines to commit to a single price for the lifetime of the ticket. This creates inefficiencies in the allocation of tickets, but also reduces price paid by consumers. I compare these results to another model in which airlines are able to price discriminate but must commit to prices in advance. In the third model, I consider a market where airlines

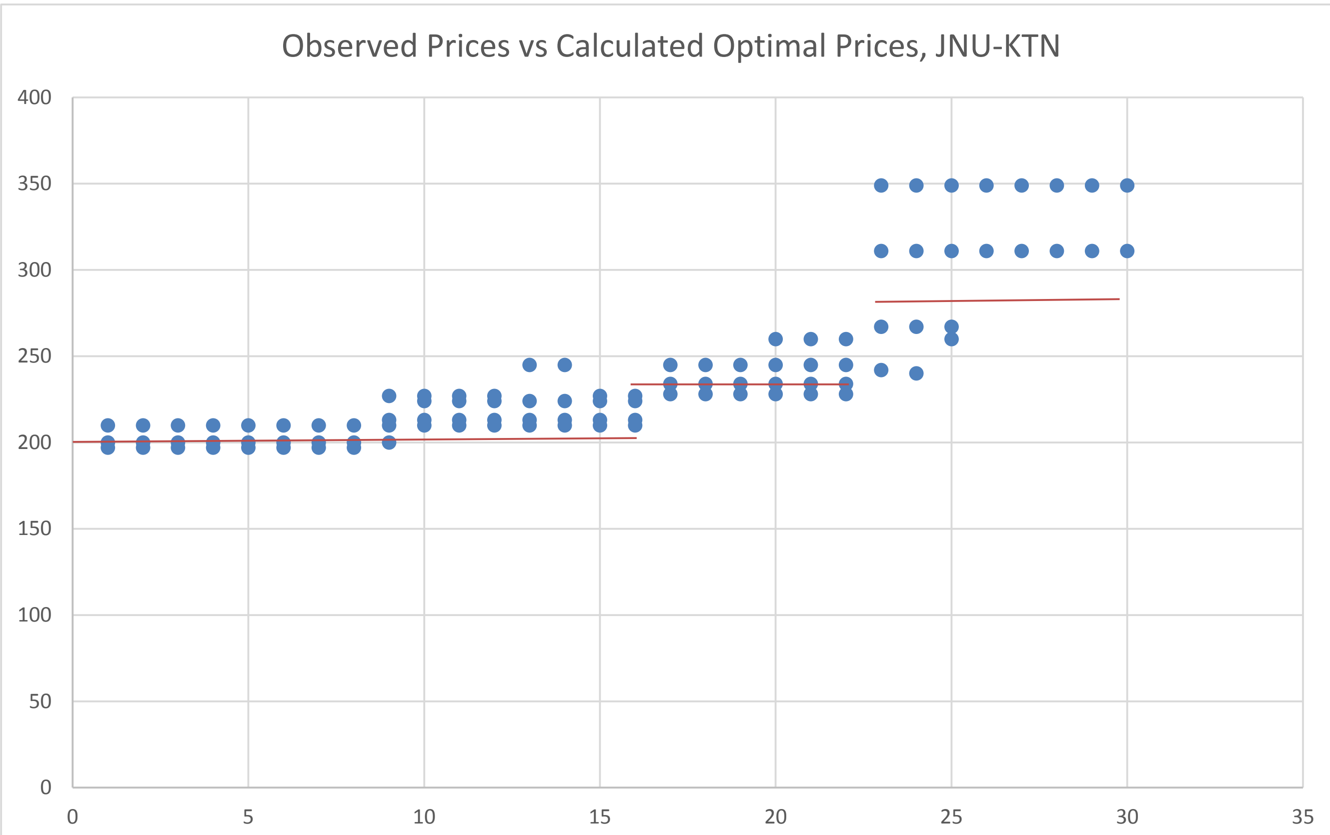

Figure 10: Predicted prices vs observed prices on JNU (Juneau, AK) to KTN (Ketchikan, AK)

have full information about consumers, which allows airlines to allocate seats more efficiently.

6.1 No price discrimination

To test the welfare effects of price discrimination, I estimate this counterfactual in which airlines commit to a single price for each flight at the beginning of the lifetime of the flight.

This pricing model has a number of effects on consumer behavior. By restricting airlines to a single price for each flight, consumers can choose to return in future periods without worrying about the prices of tickets increasing. For markets with two flights, I restrict airlines to a single price for both flights to prevent price discrimination in this way.

In this model, because airlines are restricted to setting a single price, consumers have beliefs that are consistent with that strategy. While consumers know that prices will never change, they still have beliefs about the probability of the flight selling out in each future period. To calculate the airline's optimal prices in the counterfactual, I calculate consumer beliefs about the probability of flights selling out in each period as a fixed point search. Because consumer beliefs determine their purchasing behavior, for a given price I find the beliefs that are consistent with the sell-out probabilities caused by the beliefs. Using this approach, I search for the price of tickets in each market that maximize each airline's revenue.

Under this model, airlines pick the single price to maximize revenue while facing uncertainty about the arrival rate of consumers. I find that the optimal price of tickets decreases because airlines are unable to price discriminate. The average ticket price sold was \$214.98, compared to \$227.5 under price discrimination. The average price that leisure travelers pay increases, but the average price that business travelers pay decreases. Despite the lower average price of tickets, airlines sell 5.01% fewer tickets because of their inability to lower prices when demand is low. Dynamic price adjustments are important to airlines selling to capacity, and I find that under no price discrimination, airlines are more likely to have empty seats on flights. While in my original data I found that 59% of flights sold out before the final period, only 51% of flights sell out in this model.

Because flights are less likely to sell out, the flights that do sell out early have very poor allocations. Airlines no longer save tickets to sell to business travelers near the departure date, and as a result, when flights do sell out, they are filled with leisure travelers who purchased earlier. Business travelers face lower prices but purchase fewer tickets in this counterfactual: they account for 25.8% of air travel, down from 29.2%.

Consumers perform slightly fewer searches: one incentive for consumers to search is the possibility of finding lower prices in the future. That possibility disappears in this counterfactual, while the chance for flights to sell out does not. No consumer would search more than once; the optimal strategy after arriving in the market is to pick an optimal date (dependent on their x_0 and expected probabilities of flights selling out) and return to purchase on that date. Because leisure travelers arrive very early and have a low x_0, they have a strong incentive to return at a later date. Figure 11 shows the purchase rate for leisure consumers over time. I find that the average leisure traveler performs 1.99 searches conditional on purchasing, while business travelers perform 1.08 searches. The absence of price discrimination surprisingly induces business travelers to search slightly more. The business travelers who search have low search costs and arrive early. These travelers know that price is constant under the counterfactual and they are more certain about flying in the future, whereas they would normally purchase immediately because prices are expected to increase.

Another large gain in efficiency in this counterfactual is the decrease in refunded tickets. Because prices no longer increase, leisure travelers are not pressured to purchase early. Leisure travelers purchase tickets later and as a result are less likely to pay for a ticket change. This is a large welfare increase for leisure travelers: under the status quo, 8.41% of leisure travelers that purchased a ticket ultimately realized they would not fly. Under this counterfactual, only 2.2% of consumers do.

I find that the consumer welfare increases due to decreased prices, lower search costs, and lower cancellation fees outweigh the losses from inefficient allocations. Both types of consumers have higher welfare: business travelers face lower prices, and leisure travelers have lower search costs and lower cancellation fees. Total consumer welfare is 7.2% higher in the counterfactual.

6.2 Price Commitment

My previous counterfactual considers a scenario in which airlines cannot price discriminate and must commit to a price. I estimate another alternative in which airlines have some ability to price discriminate and also must commit to a price. In this second counterfactual, airlines can set three prices for each flight (for periods 30-14, 14-7 and 7-1 days before departure) but cannot condition prices on the state of the market. The ability to change price over time allows the airline some ability to extract welfare from business consumers but does not let them adjust prices in response to high or low demand. By comparing this result to the single price case, I isolate the effects of price discrimination under price commitment.

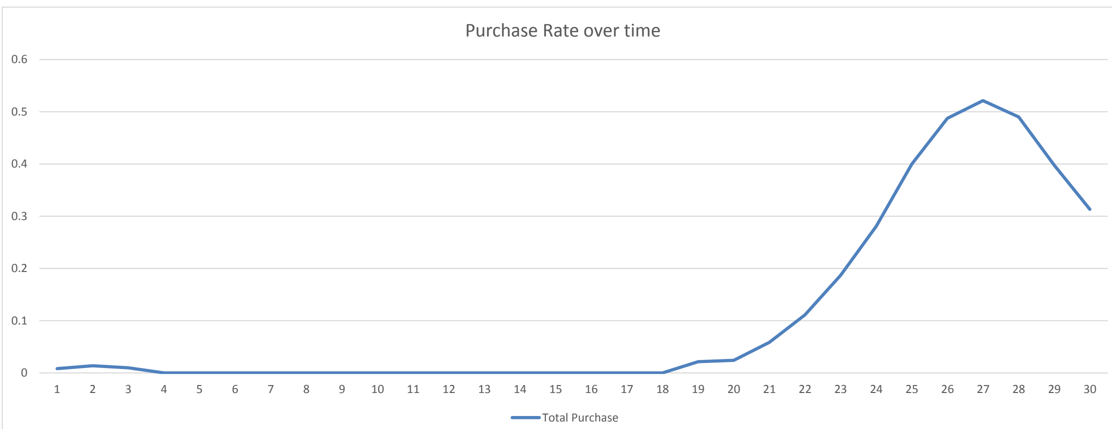

Figure 11: Average purchase rate over time in markets with a single price

In this model, airlines commit to a series of three prices for each flight. Consumers know with certainty how prices will move in the future, but not whether flights will still be available at any future date. Similar to the previous counterfactual, consumers form beliefs about when flights are more likely to sell out. In my estimation, I search for consumer beliefs that are consistent with the true probability of selling out.

I find that allowing airlines to set three prices instead of one increases airline revenue but also improves allocation efficiency. Consumer welfare decreases as a result of price discrimination, with business travelers suffering the most. Leisure travelers pay an average of $209.5 while business travelers pay $261. Compared to the single price counterfactual, leisure travelers pay $5 less, while business travelers pay significantly more. More seats are sold to high valuation business travelers and fewer flights sell out early, with business travelers purchasing 32% of tickets sold, compared to 25% in the single price counterfactual. However, the total number of tickets sold decreases again: total quantity is 6.12% lower in this model compared to the single price counterfactual and 11.13% lower than the baseline. The net result is that consumer welfare decreases significantly. One large efficiency gain in the single price counterfactual was that leisure travelers very rarely refunded a purchased ticket, because there was less pressure to purchase early. This effect disappears and I find that leisure travelers purchased earlier and more frequently paid cancelation fees. In addition, 97% of leisure travelers who chose to purchase tickets did so on the two days before price increases. As a result, the welfare of consumers in this counterfactual is 5.21% lower than the status quo and 12% lower than the no price discrimination counterfactual.

6.3 Full information

In this counterfactual, I examine the loss in total surplus caused by imperfect information. I estimate a model in which airlines have full ex-post information about the number and types of consumers who will enter the market. In this model, all consumers wait until the final day to enter, when they will realize whether they want to fly. After observing the distribution of consumers in the market, airline will set prices in the market. Airlines can always sell to capacity if they choose in this model because they face no uncertainty.

I estimate two variants of this model to compare the effects of price discrimination: one where airlines set a single price in the final period for all consumers, and one where the airlines can use third degree price discrimination to charge different prices for business and leisure travelers. I find that in the first case, airlines can under-sell tickets, whereas in the second case firms will sell out, but the allocation of tickets between groups can be inefficient.

When airlines can not third degree price discriminate, the quantity of tickets sold depends on the market. Airlines always have the option of selling all tickets on a flight, but they can also choose to undersell. In markets with low supply or high valuations of flying, airlines can choose to sell to only business travelers to keep prices high. This only occurs when capacity is large enough that business travelers cannot fill the flight, and there is a large gap between the lowest valuation business travelers and the highest valuation leisure travelers. In 9.2% of realizations, airlines choose not to sell to capacity in the final period, and when this happens, on average 27.6% of seats remain unsold. Otherwise, capacity is filled and the result is the most efficient allocation of seats: the highest valuation of consumers who still want to fly on the final date get to fly. Average price in this counterfactual is $243, with only the highest valuation consumers flying.

Despite average prices increasing, consumer welfare increases in this counterfactual. The highest valuation consumers always have access to tickets in this model because flights never sell out early. In addition, consumers never face cancellation fees and no search costs. This result is most similar to the single price counterfactual, but with the added effect of nearly efficient ticket allocation. Consumer welfare is 42.3% higher than that of the single price counterfactual and 52.5% higher than that of the base model.

Airline profit is also higher in this counterfactual. Airlines lose the revenue from cancelation fees, but the average price of tickets increases and, most significantly, the total quantity of tickets increases. The frequency of sold out flights increases from 59% to 90.8%, and total number of tickets sold increases by 18.1%. Airline's total revenue across all markets increases by 24.9%.

If we allow airlines to set separate prices for business and leisure travelers, airlines will maintain a high price for business travelers and use leisure travelers to fill capacity. As a result, I find that airlines always sell to capacity in every market, but in 11.62% of realizations airlines will sell to capacity while excluding some low valuation business travelers. Business travelers face much higher prices and leisure travelers have slightly higher prices: business travelers pay an average of $502.98 and leisure travelers pay an average of $203.44. The firm's profits increase by 37% but consumer welfare decreases by 29%.

7 Conclusion

In this paper, I explore the role of price discrimination in consumer search behavior. Using a data set of daily price and quantity observations on 55 routes, I estimate consumer demand and search behavior in the market for airfare. In my model, sophisticated consumers observe prices

and calculate the optimal period to search again for tickets based on individual preferences. My model estimation allows me to calculate the welfare effects of intertemporal price discrimination.

Through counterfactual simulation, I find that price discrimination decreases consumer welfare for both leisure and business travelers. Price discrimination's effect on consumer welfare comes from a number of sources: business travelers face higher prices; leisure travelers are pressured to purchase tickets early when they are unsure about flying, often forcing them to refund tickets; and uncertainty about future prices causes leisure travelers to search more. Price discrimination does allow a more efficient allocation of tickets, with 5.13% more tickets sold and more tickets going to high valuation consumers, but all of the efficiency gains are transferred to airlines.

In a second counterfactual, I consider a model in which airlines must commit to future prices while still able to exercise some price discrimination. I find that consumer welfare decreases as a result of inefficient allocations. Because firms cannot adjust prices to sell the optimal number of tickets, the quantity of tickets decreases.

Finally, I find that airlines' uncertainty about future demand is a large cause of inefficiency in the market. When airlines have perfect information about the quantity and preferences of consumers entering the market, there are large gains to both consumers and the airline as a result of almost perfect allocation. Within this counterfactual, I see the effects of third degree price discrimination on consumer welfare and airline profits: third degree price discrimination allows airlines to set extremely high prices for business travelers, resulting in low consumer welfare and very high profits.

Appendices

A Consumer Beliefs

Consumer beliefs in my model depend on current prices of tickets as well as the time and certain flight characteristics. However, for computational feasibility, I do not include the number of remaining seats in consumer beliefs about future price movement. To test the assumption that consumers do not base purchasing decisions on number of seats remaining, I regress the number of tickets sold in each period of my observation on the number of seats remaining as well as the variables included in my Markov process estimation: days remaining, price, price^2, price^3, days remaining, price$\times$days, 7, 14, 21 days remaining dummy variables, day of week dummy variables and

carrier fixed effects. Using multinomial logit regression, I regressed quantity of tickets (split into bins of 0, 1, 2, 3+ tickets sold) on these variables and perform a likelihood ratio test. Under the restricted model where the coefficient on number of seats remaining is 0, I get $2\ln(L_0) = -4871.2$ with 22 degrees of freedom. Under the unrestricted model I have $2\ln(L) = -4870.1$, with a test statistic of 1.098. Under a chi-squared distribution with a single degree of freedom, I cannot reject the restricted model; After accounting for prices and days remaining, I find that the additional knowledge of number of seats remaining does not help predict the number of tickets purchased by consumers.

B Identification

I estimate my model assume two populations of consumers who have an average arrival rate in the shape of a polynomial arrival function. To test alternative arrival rates, I estimate an alternative model with a single population of consumers. I find that the fit of this model has a much worse fit than my original model: the population must be willing to search for tickets (in order to capture bunching in the periods before price increases, purchase tickets in the final periods when prices are highest and purchase a steady stream of tickets over the life of the flight. I find that one population of consumers cannot capture all of the purchasing behaviors that I observed. Figure 12 shows the arrival rates of the single population, and figure 13 shows the average number of tickets purchased over time by the single population. I find that the fit of the model is significantly worse than my estimation with three populations. I am presently estimating an alternative of my model with three populations of consumers, which will appear in a future version of this paper.

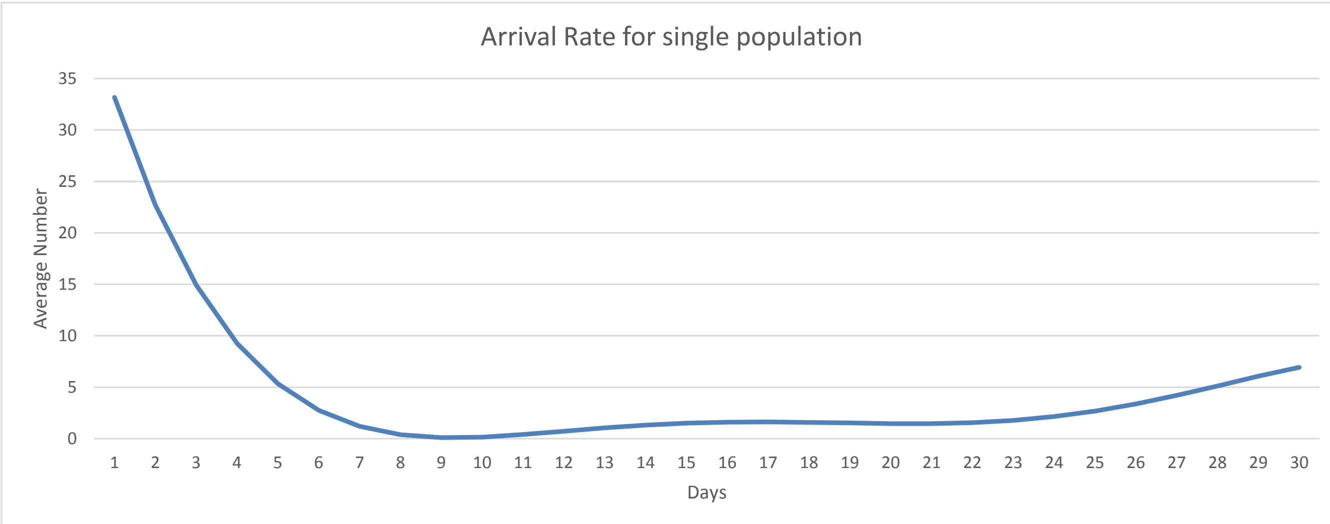

Figure 12: Average arrival rate over time in markets with a single population

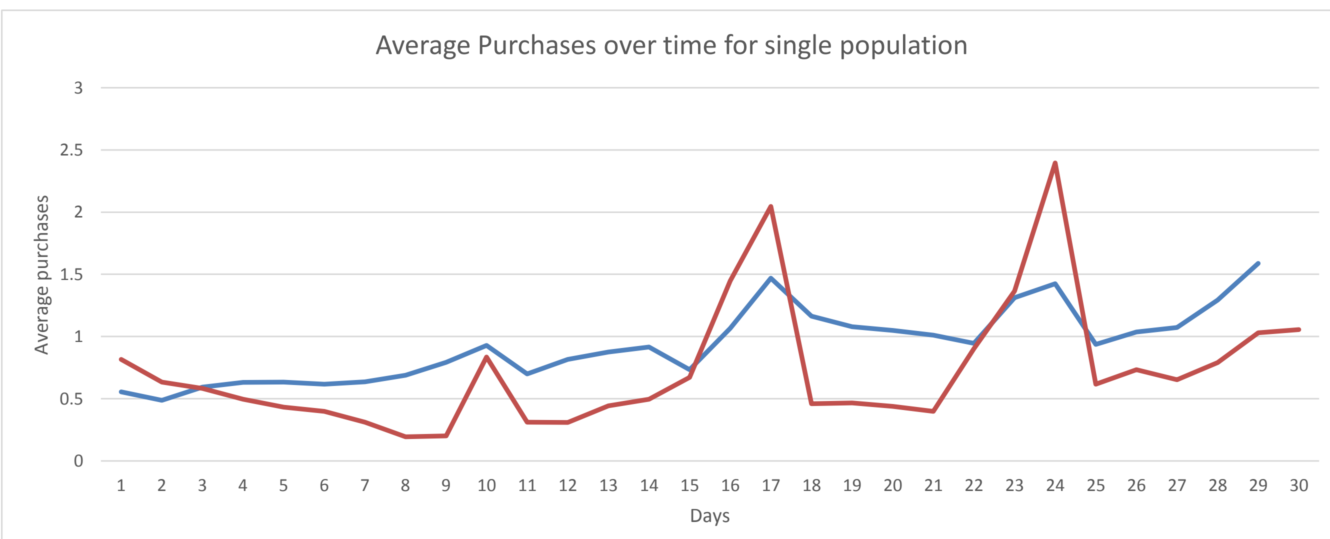

Figure 13: Average purchase rate over time in markets with a single population

References

Berry, Steven and Panle Jia (2008), "Tracing the woes: An empirical analysis of the airline industry." , National Bureau of Economic Research.

Board, Simon and Andrzej Skrzypacz (2013), "Revenue management with forward-looking buyers." *Unpublished manuscript*.

Borenstein, Severin and Nancy L Rose (1991), "Competition and price dispersion in the us airline industry." , National Bureau of Economic Research.

Borenstein, Severin and Nancy L Rose (2007), "How airline markets work... or do they? regulatory reform in the airline industry."

Chevalier, Judith A and Austan Goolsbee (2005), "Are durable goods consumers forward looking?"

Clemons, Eric K, Il-Horn Hann, and Lorin M Hitt (2002), "Price dispersion and differentiation in online travel: An empirical investigation." *Management science*, 48, 534–549.

Etzioni, Oren, Rattapoom Tuchinda, Craig A Knoblock, and Alexander Yates (2003), "To buy or not to buy: mining airfare data to minimize ticket purchase price." In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 119–128, ACM.

Gerardi, Kristopher S and Adam Hale Shapiro (2009), "Does competition reduce price dispersion? new evidence from the airline industry." *Journal of Political Economy*, 117, 1–37.

Hendel, Igal and Aviv Nevo (2004), "Intertemporal substitution and storable products." *Journal of the European Economic Association*, 2, 536–547.

Lazarev, John (2013), "The welfare effects of intertemporal price discrimination: an empirical analysis of airline pricing in us monopoly markets." *Working Paper*.

Leslie, Phillip and Alan Sorensen (2009), "The welfare effects of ticket resale." , National Bureau of Economic Research.

Nair, Harikesh (2007), "Intertemporal price discrimination with forward-looking consumers: Application to the us market for console video-games." *Quantitative Marketing and Economics*, 5, 239–292.

Williams, Kevin R (2013), "Dynamic airline pricing and seat availability." *Working Paper*.