# Pricing and travelers' decision to use frequent flyer miles: evidence from the U.S. airline industry

Bar, Michael and Chernomaz, Kirill and Diego, Escobari

San Francisco State University, San Francisco State University, The University of Texas - Pan American

28 May 2010

# Pricing and Travelers' Decision To Use Frequent Flyer Miles: Evidence From the U.S. Airline Industry

Michael Bar* Kirill Chernomaz† Diego Escobari‡

May 28, 2010

**Abstract**

Previous research on Frequent Flyer Programs (FFP) covered various topics, from analyzing the effect of international airline alliances on domestic travel demand to the effect of airport dominance and FFP on pricing. However, one important constraint in previous empirical research on FFP is the lack of a measure of these programs at a specific time-variant route and carrier level. In this chapter we use a novel way to measure the extent of FFP which allows us to analyze how these programs change from route to route, across carriers and over time. The dataset, that covers the quarters from 1993.1 to 2009.3, was constructed with data obtained from the Bureau of Transportation and Statistics and it has information on prices, proportion of frequent flyer tickets as well as various route and carrier variables. Using panel data techniques to control for unobservables along with the use of instrumental variables to control for potentially endogenous regressors, the results found are consistent with our economic model: travelers are more likely to redeem their frequent flyer miles in more expensive routes. Moreover, business travelers, who usually pay higher prices, were found to be less price sensitive than tourists when switching to buy with accumulated miles.

*Keywords*: Frequent Flyer Programs; Airlines; Panel Data

*JEL Classifications*: L11; L93; C23

*Department of Economics, San Francisco State University, kirillch@sfsu.edu

†Department of Economics, San Francisco State University, mbar@sfsu.edu

‡Department of Economics & Finance, The University of Texas - Pan American, escobari@sfsu.edu

# 1 Introduction

Since 1981 with the introduction of AAdvantage, the Frequent Flyer Program from American Airlines, Frequent Flyer Programs (FFP) have been growing enormously. It was the combination of deregulation of the industry and the introduction of computer reservations systems that gave rise to this highly popular marketing strategy. The goal is simple: to create travelers' loyalty towards a single carrier. It has been calculated that FFP have more than 80 million participants, with the three largest U.S. FFP, American Airlines' AAdvantage, United Airlines' Mileage Plus and Delta's Sky Miles, having more than 20 million members each.[1]

Despite of the large success of these programs, empirical research in economics on the analysis on FFP is scarce. The main restriction is that data on miles balances of individual travelers are unavailable to researchers. One exception is Lederman (2008), who looks at the effect of these programs on the 'hub premium' —higher fares charged by hub airlines for flights originating at the hub— by using the formation of international partnerships. Other related research has focused on airline alliances (e.g. Lederman (2007)) and partnerships (e.g. Bilotkach (2009)), but not on travelers choice on the usage of these programs. In this chapter we use a novel way to measure the extent of FFP which allows us to analyze how the extent of these programs changes from route to route, across carriers and over time.

In particular, we employ data from the Bureau of Transportation Statistics to identify how the proportion of Frequent Flyer Tickets (FFT) changes across various time-variant carrier and route characteristics. This proportion can be interpreted as an aggregation of individual travelers' decisions between paying to fly or using their accumulated frequent flyer miles to obtain a free ticket. The analysis takes advantage of a panel of carriers and routes that spans over seventeen years of data, from the first quarter of 1993 to the third quarter of 2009. Initially we analyze the market equilibrium proportion of FFT, but after instrumenting for price, we are able to measure the pricing effect on travelers' choices. We find that if average prices in a route increase by one dollar, one of every 802 paying passengers will decide to shift and use his frequent flyer miles to obtain the ticket instead of paying for it. The results also show that the lower tail of the distribution of prices has a

[1]This figure comes from www.frequentflier.com.

much larger positive effect on the proportion of FFT than the positive effect of the upper tail. This is consistent with business travelers being less price sensitive than tourists.

Our theoretical model presents an explanation of this positive effect of prices on the proportion of FFT. In a dynamic setting we show how a traveler, who faces a positive probability of making a trip, decides between paying for a ticket or using miles. When needing to fly he can either buy a ticket and accumulate miles or, if he already achieved the required number of miles, he can choose to obtain the ticket by redeeming a fixed amount of accumulated miles. The model's implication is consistent with our empirical findings: if average prices in a route are higher, the proportion of travelers who decide to use their miles is greater.

The chapter is structured as follows. Section 2 presents an overview of frequent flyer programs. Section 3 describes the empirical approach starting with the explanation of the sources of the data and sample in Section 3.1. Then we present the motivation for the empirical model in Section 3.2 and show the estimated equation, along with the selection of the instruments in Sections 3.3 and 3.4, respectively. Section 4 presents the empirical results by describing the summary statistics (4.1), explaining the results for the time trend and route specific characteristics (4.2), the price effect (4.3), airport dominance (4.4), and product quality and capacity constraints (4.5). To provide a theoretical explanation of the price effect on frequent flyer programs, Section 5 introduces a simple dynamic model of a traveler's choice between purchasing a ticket or using accumulated miles. Finally, Section 6 concludes.

## 2 Frequent Flyer Programs

One key step that led to the appearance of frequent flyer programs (FFP) in 1981 was the deregulation of airline markets that happened three years earlier. Deregulation not only allowed carriers to optimize their network structure —that led to the hub-and-spoke system which appeared shortly after deregulation— but also to offer additional incentives to attract passengers. One of those very successful ways to attract consumers is frequent flyer programs. These programs offer free travel, as the most common gift, once a customer has conducted a certain number of trips with the airline. The first airline that introduced FFP was American Airlines, but similar products were quickly introduced by United Airlines, Continental, Delta and TWA.[2]

The success of these programs has been partly attributed to the creation of consumer's loyalty towards a specific airline. The consumer has incentives to concentrate all of his business in one particular carrier because in this way it will be easier to achieve the required accumulated miles to redeem a free ticket. In addition, they are set to take advantage of a principal/agent problem when business travelers are not usually the ones who pay for their tickets, but do decide on which carrier to fly. In this way travelers will benefit from the accumulated miles, but it will be their businesses that pay for the tickets. As notes by Borenstein (1989), although this may increase the firm's costs associated with employees' travel, it also increases one non-taxed compensation received by its employees.

The effectiveness of FFP depends in large degree on the carrier's network size. Whether a carrier is able to attract travelers into their FFP is a function of its overall network size and its business size at the traveler's departing airport. Airlines who have a large network will be able to offer a larger number of alternatives for their travelers in both accumulating and redeeming the miles. This is a simple explanation of why during the nineties various airline alliances were born and also helps explain code sharing flights.[3] Likewise, the size of the carrier at a departing airport serves to attract travelers to a particular FFP. Travelers will more likely choose the dominant carrier because it has the larger set of alternatives to accumulate miles and also has the largest set of alternatives at the time of requesting a free travel. The benefits of being the dominant carrier in an airport then becomes apparent, in addition to the well known 'hub premium' —higher fares that the dominant carrier in an airport is able to charge.[4]

[2]See Mason and Baker (1996) for a history of FFP.

[3]See Lederman (2007) for international FFP partnerships and their effect on domestic airline demand and Bilotkach (2009) for local FFP partnerships.

[4]See Borenstein (1989) for the effect of route and airport dominance on fares, Lee and Luego Prado (2005) for the effect of passenger mix on the 'hub premium' and Lederman (2008) for the effect of FFP on the 'hub premium.'

# 3 Empirical Approach

## 3.1 Sources of Data and Sample

The dataset used in this chapter focuses on domestic, round-trip, coach class tickets and covers the period starting with the first quarter of 1993 and ends with the third quarter of 2009. It was obtained from the Bureau of Transportation and Statistics (BTS) website, Transtats. We use the market data and the ticket data sub-sections of the DB1B database and the segment data from the T-100. The DB1B database is a 10% random sample quarterly data of airline passenger ticket transactions. Each observation contains information about the ticket price, origin, destination and any connecting airports, number of passengers at each ticket, carrier, type of ticket (e.g. one-way, round-trip), and the service class. The T-100 segment data offers information on the total number of performed departures by carrier in an airport, as well as information on the total number of seats and transported passengers between an origin and destination airport pair.

The analysis is restricted to round-trip tickets because these tickets allow us to identify the originating airport of the ticket. Essentially we need to distinguish between SFO-DFW-SFO (San Francisco-Dallas/Forth Worth-San Francisco) and DFW-SFO-DFW because we also want to see how a carrier's specific attributes in the departing airport affect the demand for frequent flyer tickets. To restrict the analysis to economically significant routes, the sample includes all routes that had at least one carrier transporting an average of 40 passengers per week, by either direct or connecting service. The dataset is constructed in such a way that each observation in the sample corresponds to a route —a pair of origin and destination airports— served by a given carrier during a specific quarter and year. The carriers considered are AirTran, Alaska, American, Continental, Delta, Frontier, JetBlue, Northwest, Spirit, Trans World Airlines, United, and US Airways, each with its corresponding FFP partners.

Because travelers can also obtain frequent flyer miles by traveling with the carrier's FFP partner, we identified the partners of each of the carriers and considered those tickets as belonging to the main carrier. For example, passengers flying on American Eagle can count those miles towards American Airlines' frequent flyer program, hence we consider all tickets from American Eagle as if they were tickets from American Airlines. The other frequent flyer partnerships in the sample include Air Wisconsin Airlines and American West with US Airways; ATA Airlines, Atlantic Southeast Airlines, Skywest and Comair with Delta; and Mesaba with Northwest.

## 3.2 Model Setup

To be able to investigate the relationship between the usage of frequent flyer tickets and various airport, route and carrier specific characteristics, we model the total number of passengers who purchase tickets using frequent flyer miles, $Y_{ijt}$, as a binomial random variable, $Y_{ijt} \sim Bin\left(n_{ijt}, \pi_{ijt}\right)$. The subscript $i$ refers to the carrier, $j$ refers to the departure and destination airport pair, and $t$ refers to the time period. Therefore, the total number of travelers served by carrier $i$ in route $j$ during time $t$ is given by $n_{ijt}$, with $\pi_{ijt}$ being the probability that a given traveler on $ijt$ obtains the tickets through a frequent flyer program.[5] Thus, the proportion of travelers with free ticket in $ijt$ is $P_{ijt} = Y_{ijt}/n_{ijt}$ with $E\left(P_{ijt}\right) = \pi_{ijt}$ and $Var\left(P_{ijt}\right) = \pi_{ijt}\left(1 - \pi_{ijt}\right)/n_{ijt}$. Notice that the variance of the proportion is decreasing with the number of travelers on the route, i.e. $n_{ijt}$.

The object of our study is the proportion of free tickets in each route in the population, i.e. $\pi_{ijt}$. We model it as a function of carrier, airport and route characteristics that are allowed to change over time, $\pi_{ijt} = F\left(\mathbf{X}\beta\right)$. $\beta$ is the vector of the coefficients of interest that we will estimate, $\mathbf{X}$ is the matrix of carrier, airport and route characteristics and $F$ is a monotonically increasing function that maps the value of characteristics into the $[0, 1]$ interval. The empirical model can then be written as:

$$P_{ijt} = F\left(\mathbf{X}\beta\right) + \varepsilon_{ijt} \tag{1}$$

where $E\left(\varepsilon_{ijt}\right) = 0$ and $Var\left(\varepsilon_{ijt}\right) = F\left(\mathbf{X}\beta\right)\left[1 - F\left(\mathbf{X}\beta\right)\right]/n_{ijt}$. To obtain a linear model, we apply the inverse transform $F^{-1}$ to both sides of Equation 1 and obtain a Tailor series approximation around $\varepsilon_{ijt} = 0$. Thus, the linear model can be written as:

$$F^{-1}\left(P_{ijt}\right) = \mathbf{X}\beta + u_{ijt} \tag{2}$$

where $u_{ijt} = \varepsilon_{ijt}/F'\left(\mathbf{X}\beta\right)$. To estimate Equation 2 we need to specify the function $F^{-1}$. As a first approach we will use a linear probability model where $F^{-1}\left(P_{ijt}\right) = P_{ijt}$. OLS

[5] It is straightforward to see that this binomial distribution has mean $E\left(Y_{ijt}\right) = n_{ijt}\pi_{ijt}$ and variance $Var\left(Y_{ijt}\right) = n_{ijt}\pi_{ijt}\left(1 - \pi_{ijt}\right)$.

estimates with appropriate weights to take care of the heteroskedasticity present in the model will be unbiased and consistent, but the predicted values of $P_{ijt}$ are not restricted to lie on the $[0, 1]$ interval.[6] To take care of this problem we will provide estimates using two functional forms for $F^{-1}$. The first one is the log-odds ratio:

$$F^{-1}\left(P_{ijt}\right) = \log\left(\frac{P_{ijt}}{1 - P_{ijt}}\right) \tag{3}$$

which gives us the logistic function $F\left(\mathbf{X}\beta\right) = \exp\left(\mathbf{X}\beta\right) / \left(1 + \exp\left(\mathbf{X}\beta\right)\right)$ that can be estimated via maximum likelihood. When this transformation is estimated via OLS and not via maximum likelihood, it requires all observation $P_{ijt}$ to be strictly between 0 and 1. To be able to use panel data techniques that consider different structures in the error term and allow us to have values of $P_{ijt}$ of 0 and 1, we will use the second functional form for $F^{-1}$. This one was proposed by Cox (1970) and it is given by:

$$F^{-1}\left(P_{ijt}\right) = \log\left(\frac{P_{ijt} + a_{ijt}}{1 - P_{ijt} + a_{ijt}}\right) \tag{4}$$

where $a_{ijt} = \left(2n_{ijt}\right)^{-1}$.

### 3.3 Estimated Equation and Panel Structure

We now rewrite Equation 2 to emphasize the fact that we are considering the proportion of frequent flyer tickets, $P_{ijt} = PROPFFT_{ijt}$, as a function of various airport, route, and carrier specific characteristics taking into account the panel structure of the data. Hence, we break down the error term into different components, $u_{ijt} = \upsilon_t + \eta_{ij} + \mu_{ijt}$. The resulting reduced-form equation is:

$$\begin{aligned} F^{-1}(PROPFFT_{ijt}) = {} & \beta_1 MEANFARE_{ijt} + \beta_2 MILES_j + \beta_3 DEPAHUB_{ij} \\ & + \beta_4 PROPDEST_{ijt} + \beta_5 PROPDEPA_{ijt} + \beta_6 PROPDIRECT_{ijt} \\ & + \beta_7 LOADFACT_{ijt} + \upsilon_t + \eta_{ij} + \mu_{ijt} \end{aligned} \tag{5}$$

Let $\upsilon_t$ denote any unobservervable time specific effect, $\eta_{ij}$ denote the unobservable carrier/route time-invariant specific effect and $\mu_{ijt}$ denote the remaining disturbance. The dependent variable is our measure of frequent flyer tickets, with $PROPFFT$ calculated as

[6] The weights will also make sure that each observation is given the appropriate importance according to the corresponding total number of travelers.

the ratio of frequent flyer tickets to the total number of tickets. The number of frequent flyer tickets is obtained from the number of tickets with a price equal to zero, as recorded in the DB1B database. $MEANFARE$ is the average price paid by passengers flying with carrier $i$ during quarter $t$ in route $j$. Notice that route $j$ is defined as a combination of origin and destination airports, so the average is taken over all direct and connecting flights. Two route specific characteristics that we have are $MILES$, that is the number of miles between the origin and destination airports and $DEPAHUB$, that is a dummy variable that equals to one when the departing airport $j$ is a hub for carrier $i$ and zero otherwise. Notice that we will only be able to identify the route specific characteristics if Equation 5 is estimated without route specific fixed effects.

In addition to the main variable of interest, $MEANFARE$, and the two time-invariant route/carrier characteristics, we constructed four additional variables that change across carriers, routes and over time. The first two are similar to $DEPAHUB$ and should capture the effect of time-variant airport presence or airport dominance on frequent flyer miles usage. These variables are included to evaluate the idea that passengers will typically join the program of the dominant carrier at their departing airport. The reason is simple, passengers will be able to accumulate miles more easily if they join the program of the dominant carrier at their departing airport because most of their destinations will be covered by this carrier. Moreover, the dominant carrier will also give them more options in terms of destinations at the time of redeeming their miles. The variable $PROPDEST$ is the proportion of nonstop destinations. It is obtained by dividing the nonstop destinations of carrier $i$ out of the departing airport in route $j$ during time $t$ by all the nonstop destinations out of route $j$'s departing airport during time $t$. Likewise, $PROPDEPA$ measures the proportion of departures out of the departing airport in route $j$ that belong to carrier $i$. According to an airport dominance story, both, more destinations and more departures from a specific carrier in an airport should attract more people to its frequent flyer program, hence $PROPDEST$ and $PROPDEPA$ would have a positive effect on $PROPFFT$.

The variable $PROPDIRECT$ —proportion of carrier $i$'s direct flights on route $j$ during time $t$— was obtained by dividing carrier $i$'s total number of direct flights by the total number of flights offered by the same carrier. $LOADFACT$, the average load factor or capacity utilization, is a measure of the usage of aircraft's capacity. Its value is in the $[0, 1]$

interval, being equal to one when aircrafts operate at full capacity.[7]

Equation 5 will be estimated with different specifications of the error term. When focusing on the regressors that change across carriers, routes and over time, we will include two sets of fixed effects. First, we use time fixed effects to control for changes over time in the industry-level demand, which may be correlated with the industry-level trends, such as the adoption of new technologies like Internet bookings, but that affect all carriers and all routes equally. Second, we use route/carrier fixed effects to control for time-invariant unobservable factors that affect frequent flyer tickets' demand. Because each departing airport belongs to a specific route, these fixed effects will control for the hub effect or level of dominance of an airport. Time-invariant regressors (e.g. $DEPAHUB$) cannot be separately identified when these fixed effects are included, so we will be able to see how the other two measures of airport dominance, $PROPDEST$ and $PROPDEPA$, affect the proportion of frequent flyer tickets once the time-invariant component of airport dominance is controlled for. Notice that because $DEPAHUB$ is a dummy equal to one when the departing airport is a hub, it assumes that the effect should be equal across different hubs from different carriers. However, including carrier-route fixed effects controls for the time-invariant hub effect without having to impose a one-size-fits-all effect for airport dominance.

### 3.4 Instruments

One concern in the estimation of Equation 5 is that fares and the allocation of frequent flyer tickets may be determined simultaneously, making fares correlated with the error term. For example, Escobari (2009) empirically shows that carriers will be setting higher prices during *ex-ante* known peak periods. In this case it is likely that during high demand periods carriers will also restrict the availability of tickets assigned to passengers that obtain them through frequent flyer miles. We expect that this effect is captured by capacity utilization, $LOADFACT$, as well as by our various time and route/carrier fixed effects. However, as mentioned by Lederman (2007), there may be factors such as advertisement of a particular frequent flyer program, that can affect prices as well as the proportion of people choosing to enroll and fly using accumulated miles. Here $LOADFACT$ and the fixed effects variables may not be enough, hence we need to instrument for the potential endogeneity of fares.

[7]A more detailed construction of the variables is presented in Appendix A.

The selection and construction of the instruments is similar to the ones used in Lederman (2007) and come from the discrete-choice demand literature. The idea is that firms that offer multiple products, such as connecting and direct flights in airlines, will jointly set the prices for these products. There is a degree of substitutability between direct and connecting service because as a given carrier increases its prices for direct service, some of its consumers may not only shift to its competitors but also to its own connecting flights. The dataset DB1B allows us to distinguish between these two products within the same airport pair, so we use this to construct two instruments. The first one is the dummy variable $CONNECT$, which takes the value of one when carrier $i$ offers connecting service in route $j$ and zero otherwise. The second variable, $NUMCONN$, counts the total number of connecting combinations that carrier $i$ offers in route $j$.

# 4 Empirical Results

## 4.1 Summary Statistics

The summary statistics of the variables is presented in Table 1. The mean of the proportion of frequent flyer tickets indicates than on average airlines have 4.85% of the travelers using their accumulated miles to obtain a ticket. To measure prices at the carrier, route and quarter level we use three variables. The average fare ($MEANFARE$) —which excludes free tickets— has an average of US$ 180.9 and the $20^{th}$ and $80^{th}$ percentiles of fares ($20PCTFARE$ and $80PCTFARE$) have an average of US$ 113.7 and US$ 238.4, respectively. The time-invariant variables $MILES$ and $DEPAHUB$ show that the nonstop distance between airports in a route ranges from 67 to 6,089 miles and that on average about 20.4% of the observations have the carrier's hub as the departing airport. The other four controls as well as the two instruments for the fare variables complete the table. The sample used is an unbalanced panel with 45,000 different origin and destination pairs for a total of 474,856 airline-route-carrier observations.

[Table 1, here]

### 4.2 Trend and Route Specific Characteristics

The results from the estimation of Equation 5 using the linear transformation, $F^{-1}(P_{ijt}) = P_{ijt}$, are presented in Table 2. All estimates in this table were obtained using the total number of observations per route-carrier-time combination as weights to account for potential heterosckedaticity and to give each observation the appropriate importance. The sample starts with the first quarter of 1993 until the third quarter of 2009 and the numbers in parentheses are heteroskedasticity robust standard errors, correction included in addition to the weights. In different columns we provide different specifications for the error term, starting with the first column that gives the OLS estimates from pooling across carriers, routes, and time periods and the second column that provides the estimates for the variance component —random effects— model. Both of these first two specifications allow us to identify the effect of time-invariant route/carrier specific characteristics as well as the existence of a time trend. Consistent with the existence of an airport dominance effect, the proportion of frequent flyer tickets is larger for the carrier that has a hub at the departing airport. The statistically significant point estimate of 0.637 in the random effects specification indicates that when the departing airport is a hub, the proportion of frequent flyer tickets increases by 0.637 percentage points. This is an economically significant effect given the average of the proportion of frequent flyer tickets of 4.85%. In addition, the positive coefficient in $MILES$ is consistent with travelers picking to redeem their miles in longer haul routes.

[Table 2, here]

The evolution of the proportion of frequent flyer tickets can be observed by looking at the variable $YEAR$, with the negative coefficient showing that the proportion of frequent flyer tickets has been decreasing over time. The third column, that further controls for unobserved route and carrier specific characteristics, provides additional support for this negative effect.

To complete the set of estimates for the trend and the route specific characteristics, the maximum likelihood estimates using the transformation presented in Equation 3 are presented in Table 3. Moreover, the pooled, random effects and the route/carrier fixed effects specifications using the Cox transformation shown in Equation 4 are presented in the first, second and third columns of Table 4, respectively. The coefficients are all statistically significant and the signs are the same as the ones obtained in Table 2 and discussed above.

[Table 3, here]

[Table 4, here]

### 4.3 The Price Effect

The main variable of interest is the price. We are interested in knowing how average prices affects the equilibrium proportion of frequent flyer tickets as well as how travelers respond to prices. Table 2 presents the results for the linear transformation, ignoring the potential endogeneity of $MEANFARE$. All four columns have very similar positive and highly statistically significant coefficients. We focus on the last column, as it is the one that controls for unobserved route/carrier characteristics as well as any unobserved time effects. The coefficient of 0.0393 indicates that an increase of one standard deviation in $MEANFARE$ —an increase of US$ 65.76— increases the proportion of FFT by 2.58 percentage points.[8] This figure corresponds to a 0.39 standard deviations increase in FFT, which is economically significant.

One limitation of the linear model discussed above is that it does not restrict the fitted values of the proportion to be between zero and one. To overcome this restriction, we present additional estimates in Tables 3 and 4 that employ the nonlinear transformations presented in Equations 3 and 4, respectively. The MLE estimates in Table 3 were obtained with time-specific effects, robust standard errors and without instrumenting for any of the fare variables. For this specific transformation the marginal effect of the regressors is given by $\partial\pi_{ijt}/\partial x_{ijt,k} = \beta_k\pi_{ijt}(1-\pi_{ijt})$, where $\pi_{ijt} = E(PROPFFT_{ijt})$. To calculate this marginal effect we will use the sample average of $PROPFFT_{ijt}$ as $\pi_{ijt}$, which is 0.0485, and it is presented in Table 1. Then from the $MEANFARE$ coefficient in the first column of Table 3 we can read that a one standard deviation increase in mean fares increases the proportion of FFT by 0.616 percentage points, which corresponds to a 0.09 standard deviations increase in FFT. Table 4 shows the estimates using the Cox transformation presented

[8]This is calculated as $(\partial PROPFFT \ / \ \partial MEANFARE) \times 100 = (0.0393 \times 65.76) \times 100$. A one dollar increase in mean fares, will increase the proportion of FFT by 0.0393 percentage points.

in Equation 4 for various specifications of the error term; pooling across panels, random effects and two specifications with fixed effects. For a large number of observations within each carrier, route, and time, $n_{ijt}$, the marginal effect of a regressor can be approximated by $\partial\pi_{ijt}/\partial x_{ijt,k} \approx \beta_k \pi_{ijt}(1-\pi_{ijt})$, where $\pi_{ijt} = E(PROPFFT_{ijt})$.[9] Column 4, our preferred specification because it controls for most of the unobservables, indicates that a one standard deviation increase in mean fares increases the proportion of FFT by 2.53 percentage points, figure that is very close to the 2.58 point estimate of the linear model.

The point estimates of the price effect presented in Tables 2 through 4 can be interpreted as the price effect on the equilibrium proportion of FFT. Then, the positive coefficient can have both, supply and demand side explanations. For example, on the demand side it could be travelers using frequent flyer tickets in more expensive routes. Moreover, on the supply side it could be carriers restricting the number of available frequent flyer tickets while setting lower fares. The concern is that prices could be correlated with some supply side unobservables, making prices endogenous in the estimation of Equation 5. To be able to restrict the interpretation to a demand side story, we proceed by instrumenting for the price variable and provide two stage least squares estimates (2SLS) of Equation 5. The first stage estimates using the instruments and the matrix of regressors $\mathbf{X}$ are presented in Table 6 of Appendix B.[10] This first stage is the pricing equation as specified in Equation 7 with the dependent variable being the mean fare for column one and the $20^{th}$ and $80^{th}$ percentiles for the second and third columns. These estimations use the full sets of route/carrier fixed effects as well as the time fixed effects. The second stage coefficients are obtained using the Cox transformation of Equation 4 and are presented in Table 5. Interestingly, the highly

[9]To derive the marginal effect when using the transformation in Equation 4, we first set it equal to $\mathbf{X}\beta$ and isolate $\pi_{ijt}$ to obtain:

$$\pi_{ijt} = F(\mathbf{X}\beta) + \frac{(1+a_{ijt})\exp(\mathbf{X}\beta) - a_{ijt}}{1+\exp(\mathbf{X}\beta)}$$

Then, the marginal effect of $x_{ijt,k}$ on $\pi_{ijt}$ is:

$$\frac{\partial \pi_{ijt}}{\partial x_{ijt,k}} = \beta_k \frac{\exp(\mathbf{X}\beta)}{1+\exp(\mathbf{X}\beta)} \left[\frac{1+2a_{ijt}}{1+\exp(\mathbf{X}\beta)}\right]$$

For large $n_{ijt}$, the marginal effect is approximately: $\partial\pi_{ijt}/\partial x_{ijt,k} \approx \beta_k \pi_{ijt}(1-\pi_{ijt})$.

[10]The first stage estimates in Appendix B also provide weak instruments test using the conventional F statistic on the set of excluded instruments of the second stage. The F statistics of more than 10 in all three specifications signals that we do not have a weak instruments problem.

significant $MEANFARE$ estimated coefficient of 0.027 in column one is larger than the 0.008 of the fourth column in Table 4, obtained when not instrumenting for fares. This 2SLS estimate indicates that a one standard deviation increase in mean fares increases the proportion of FFT by 8.19 percentage points, which corresponds to a 1.25 standard deviations increase in the proportion of FFT. In dollar terms, a one dollar increase in average fares increases the proportion of FFT by 0.12 percentage points. In other words, if on a given route a carrier transports 802 paying passengers per week, a one dollar increase in fares will make one of those passengers decide to use his frequent miles to obtain the ticket. Given the instrumentation of the fare variable, this effect has only a demand side interpretation.

The intuition behind this positive coefficient is the following, passenger enrolled in frequent flyer programs can accumulate miles as they travel or through various other channels, such as AAdvantage.[11] Once the traveler has accumulated a certain amount of miles, let's say more than 25,000 miles, he will be able to redeem a fixed amount of miles to obtain a free ticket. When the traveler needs to fly again and he already has enough miles to obtain a free ticket, he needs to decide whether to buy the ticket and keep accumulating miles or use his accumulated miles and fly for free. As we argue in Equation 5, such decision depends on various carrier, route and time characteristics, and in particular, it depends on the average price of the ticket. If the ticket is more expensive, he will be more likely to choose to obtain it using his accumulated miles, but if it is cheaper, he will be more likely to choose to pay for it. That is exactly what the positive coefficient of prices means. As average fare increases, the proportion of travelers who choose to fly using their accumulated miles also increases. The theoretical model presented in Section 5 formalizes this idea to show how in a dynamic model the proportion of FFT is larger when average prices are higher.

[Table 5, here]

Because it has been widely documented that there is significant price dispersion in tickets bought within the same route (see for example Borenstein and Rose (1994) or

[11] AAdvantage is the credit card associated with American Airlines, and this one offers frequent flyer miles for purchases made with the card.

more recently Gerardi and Shapiro (2009)), investigating just the effect of the average route/carrier prices is restrictive. Therefore, we extend the analysis to see how the upper and the lower tails of the price distributions affect the proportion of FFT. To do this we constructed two additional variables, the $20^{th}$ and the $80^{th}$ percentiles of fares paid. Columns two and three of Table 3 present the maximum likelihood estimates when using the log-odds ratio transformation presented in Equation 3 and columns two and three of Table 5 present the estimates from the 2SLS using the Cox transformation of Equation 4. While the MLE estimates do not control for the potential endogeneity of the fares variable, the 2SLS estimates do. Interestingly, we observe that all fare coefficients of the different specifications of Table 5 are highly statistically significant. The magnitude of the coefficients indicates that more expensive fares have a smaller effect on the proportion of FFT than less expensive fares. The coefficient of the $20^{th}$ percentile is a little more than double of the mean fare coefficient and coefficient of the $80^{th}$ percentile is a little less than half of the mean fare coefficient. The differences in the coefficients can be explained by the fact that more expensive tickets are usually obtained by business travelers, who are less sensitive to price changes. On the other hand, cheaper tickets are usually bought by more price sensitive travelers, tourists. Then it makes sense for business travelers who buy in the upper tail of the price distribution to respond less and for more price sensitive buyers in the lower tail to respond more.

### 4.4 Airport Dominance

Airport dominance, as discussed in Borenstein (1989), refers to a particularly large share of passengers that a specific airline may have in an airport. The well documented effect of airport dominance is that the dominant carrier in an airport will be able to charge significantly higher prices than the rest of the carrier serving that airport. Because dominance at an airport is related to the hub-and-spoke network of a carrier, this is also referred to as the 'hub premium.' The existence of a hub premium is important in our estimation of the effect of prices on the proportion of FFT because airport dominance is related to the enrollment of a particular frequent flyer program. Travelers are more likely to enroll in the program of the dominant carrier in their departing airport, hence not appropriately controlling for airport dominance will bias our estimates of the price effect. In particular, a positive correlation between prices and airport dominance and a positive correlation of FFT usage and airport dominance will bias our estimate of prices upwards, overestimating the price effect.

In section 4.2 we discussed how when route/carrier fixed effect are not included, we are able to identify a positive effect of $DEPAHUB$. However, the specifications in the fourth column of Table 4 and all specifications in Table 5, with this set of fixed effects, are aimed at controlling for time-invariant characteristics, which wipes out any time-invariant hub effect. Then, the concern is whether there exists any remaining airport dominance effect that changes over time. To capture this we included the variables $PROPDEST$ and $PROPDEPA$ that measure the relative importance of a given carrier in an airport in terms of the number of destinations and number of departures, respectively. The last column in Table 4 shows a positive sign for both coefficients, consistent with what we expect from airport dominance. When instrumenting for mean fares, the first column of Table 5 shows that only $PROPDEPA$ has a positive and significant effect. An increase of 10 percentage points in the proportion of departures of a carrier out of the departing airport increases its proportion of FFT by 0.24 percentage points.

### 4.5 Product Quality and Capacity Constraints

The two last controls included in Equation 5 are the proportion of direct flights on the route, $PROPDIRECT$, and the average aircraft capacity utilization or load factor, $LOADFACT$. Because carriers offer both direct and indirect service between the city pairs on a route, $PROPDIRECT$ is used to see if carriers with a larger proportion of passengers serviced in direct flights is associated with larger or smaller FFT usage. This variable can be viewed as a measure of quality because non-stop or direct service is usually regarded of a higher quality than indirect service. The negative and highly significant coefficient of $PROPDIRECT$ in all the specifications across all tables suggests that carriers that serve routes with a larger proportion of non-stop service —higher quality— have a lower proportion of FFT. From the results in the first column of Table 5, we can say that a 10 percentage points increase in the proportion of direct flights in a route decreases the proportion of FFT by 0.35 percentage points.

The variable $LOADFACT$ is included in the estimation of Equation 5 to capture the role of capacity constraints. We were not able to obtain a sign and a coefficient robust to different specifications. When not instrumenting for fares, the regression results show a highly significant negative coefficient, but when instrumenting for fares the negative effect disappears. The negative sign would be consistent with carriers allocating fewer frequent flyer tickets in more congested routes, where capacity constraints make paying passengers more valuable to the carrier than passengers that travel with frequent flyer miles. An important idea behind free tickets is that carriers accommodate those travelers when capacity constraints are not binding, hence the costs for the carrier associated with those tickets are expected to be lower.

# 5 A Simple Model of the Price Effect

In this section we present a simple dynamic model that motivates the findings that the proportion of tickets obtained through frequent flyer miles is larger when the average price is higher. We model the decision of an individual who every time he needs to fly, has to decide whether to pay for the ticket or, if he has enough accumulated miles, to obtain the ticket with his miles. Let $n$ be the time (in months) since he last flew and the number of accumulated miles in his frequent flyer account be given by $k$. Each period the individual will fly with probability $\pi$ and stay home with no activity in his frequent flyer account with probability $1 - \pi$. In the case where the individual needs to make a trip, this trip is characterized by a pair of variables: distance between the airports and price, $(d, p)$. The price $p$ is only paid if the individual decides not no use his miles. In this scenario, he will be able to save $p$ dollars and next period his number of accumulated miles will increase to $k' = k + d$. In the event he decides to use his miles, he will need to exchange $a$ number of miles to obtain a free ticket, then he saves $p$, and the following month he will be left with $k - a$ miles. Notice that in any given period he will only be able to use his miles if $k > a$.[12]

An important characteristics in the way frequent flyer programs work is that miles expire after certain amount of time of account inactivity.[13] To be able to model the expiration

[12]This threshold $a$ varies from carrier to carrier. The most typical value is 25,000 miles for round trip tickets during peak demand periods.

[13]For example, the AAdvantage account from American Airlines expire after 18 months of inactivity, the Dividend Miles program from US Airways has its miles expire after 18 months as well as Mileage Plus from

of miles, we count the number of months since last account activity. If there is activity this period, the expiration clock is reset to zero, so next period $n' = 1$. In case there is no activity this period, miles do not change, but their age increases, $n' = n + 1$. With $\delta$ being the time discount factor, the dynamic decision problem of a traveler can then be described with the following Bellman's equation:

$$V(k,n) = \pi \max\{p + \delta EV(k-a,1), \delta EV(k+d,1)\} + (1-\pi)\delta EV(k,n+1)$$

If the account does not have any activity for certain amount of time $\bar{n}$, then miles expire. In that case the value function is defined by:

$$V(k,n) = V(0,0) \qquad \forall n \geq \bar{n}$$

When the individual has enough miles, $k > a$, he will use them for the current flight if and only if:

$$\begin{aligned} p + \delta EV(k-a,1) &\geq \delta EV(k+d,1) \\ p &\geq \delta E[V(k+d,1) - V(k-a,1)] > 0 \end{aligned}$$

In other words, for any price above $p_{min} = \delta E[V(k+d,1) - V(k-a,1)]$, the individual will use his accumulated miles to obtain a ticket. The difference $V(k+d,1) - V(k-a,1)$ represents the extra future utility from purchasing the ticket with money, and thereby increasing the stock of miles in the next period. Only if the flight is expensive enough, the traveler would prefer to use up existing miles. Moreover, when the discount factor $\delta$ is greater, the minimum price that induces travelers to use their miles increases because the future gain of accumulating miles is more valuable.

The term on the right-hand side is strictly positive since $V$ is strictly increasing in the first argument. This proves that the average price among passengers who used miles is higher than the unconditional average ticket price:

$$E[p|p \geq p_{min}] > E(p) \tag{6}$$

Hence, when the average price in a route is higher, the proportion of travelers who decide to use their miles is greater.

United.

# 6 Conclusion

This chapter sets to show the importance of pricing in the usage of Frequent Flyer Tickets (FFT). The theory section presents a simple dynamic model that illustrates how prices affect a traveler's decision between paying or using his accumulated miles to obtain a ticket. The model's empirical implication is that when average prices are higher, there is a larger proportion of travelers who use their accumulated miles to obtain a free ticket.

The empirical section models the same traveler's decision and uses aggregate data from the Bureau of Transportation and Statistics to estimate how prices and other route and carrier characteristics affect the proportion of travelers who fly using FFT. Initial estimations focus on the equilibrium number of FFT and employ various specifications of the error term to control for carrier and route time-invariant unobserved specific characteristics as well as unobserved time-variant characteristics common to all carriers and routes. OLS, MLE, FE and RE estimates combined with linear and nonlinear transformations of the proportion of frequent flyer tickets all consistently found a positive correlation between prices and FFT usage. Moreover, weighted 2SLS estimates that account for potential heteroskedastic errors and the endogeneity of fares identified the travelers response to changes in average prices. The results were found to be consistent with the theoretical model's implications, namely, higher average prices increase the proportion of passengers who use their free FFP tickets to fly. In addition to the response to average prices, we also showed the effect of the upper and lower tails of the pricing distribution. The response to the upper tail was smaller than the response of the lower tail, indicating that business travelers, who usually pay higher fares than tourists, are less likely to switch to FFT in response to a price increase.

The results also show that the proportion of FFT decreased over time, is larger in longer haul routes, and increases when the carrier has a hub at the departing airport. Our time-variant airport dominance regressors were found to have a significant effect on the proportion of FFT, even after controlling for any unobservable time-invariant hub effect. Other time-variant covariates indicate that carriers that serve routes with a larger proportion of non-stop service also have a lower proportion of FFT. Finally, the results showed some evidence that carriers restrict the number of passengers flying with free tickets in more congested routes.

# A Variable Construction

Variables were constructed excluding all tickets priced below $20 and all tickets that had questionable fare values based on credit limits, as classified by the BTS. The description of the variables used in this chapter is the following:

$PROPFFT_{ijt}$: Number of tickets priced at zero divided by the total number of tickets that belong to carrier $i$ on route $j$ during quarter $t$. *Source*: DB1B.

$MEANFARE_{ijt}$: Average price paid by all passengers traveling on the observed airline $i$ during quarter $t$ on the observed combination of origin and destination airports in route $j$. The average is taken from all passengers flying with direct or connecting service and fares are represented as oneway fares by dividing round trip ticket prices by two. *Source*: DB1B.

$20PCTFARE_{ijt}$: The $20^{th}$ percentile fare paid by passengers from carrier $i$ on route $j$ during quarter $t$. *Source*: DB1B.

$80PCTFARE_{ijt}$: The $80^{th}$ percentile fare paid by passengers from carrier $i$ on route $j$ during quarter $t$. *Source*: DB1B.

$MILES_j$: Number of nonstop miles between the origin and destination airports in route $j$. *Source*: DB1B.

$DEPAHUB_{ij}$: A dummy variable equal to one if carrier $i$ has a hub at the departing airport corresponding to route $j$.

$PROPDEST_{ijt}$: Proportion of nonstop destinations from the originating airport in route $j$ that belong to carrier $i$. *Source*: DB1B.

$PROPDEPA_{ijt}$: Proportion of departures from the originating airport in route $j$ that belong to carrier $i$. *Source*: T-100 Domestic Segment.

$PROPDIRECT_{ijt}$: Proportion of direct flights, constructed as the number of direct flights offered by carrier $i$ on route $j$ divided by the total number of flights (direct and connecting) serving route $j$. *Source*: DB1B.

$LOADFACT_{ijt}$: Load factor in a segment is defined as the total number of passengers divided by the total number of seats. Because we have routes $j$ that involve more than one segment and different combinations of segments for the same carrier $i$ (e.g. SFO-MIA, SFO-DFW-MIA or SFO-LAX-MIA), we calculate the load factor in a route as the weighted average of load factors in each of the segments. The weight is the trafic of passengers flying on the speficic route. *Source*: T-100 Domestic Segment and DB1B.

$CONNECT_{ijt}$: A dummy variable equal to one if the carrier $i$ offers connecting service in route $j$, zero otherwise. (instrument) *Source*: DB1B.

$NUMCONN_{ijt}$: Count on the total number of connecting combinations that carrier $i$ offers in route $j$. (instrument) *Source*: DB1B.

# B First Stage Regressions

The first stage regression for the 2SLS estimation of Equation 5 is given by:

$$MEANFARE_{ijt} = \alpha_1 CONNECT_{ijt} + \alpha_2 NUMCONN_{ijt} + \mathbf{X}\delta + \varepsilon_{ij} + \mu_{ijt} \quad (7)$$

where the dependent variable is either $MEANFARE$, $20PCTFARE$, or $80PCTFARE$. $CONNECT$ and $NUMCONN$ are the excluded instruments for the fare variable, and the matrix $\mathbf{X}$ is the same as the one defined for Equation 5. From the F tests reported in Table 6, with the corresponding p-values, we can see that the instruments comply with the identification assumption of being correlated with fare.

[Table 6, here]

# References

Bilotkach, V., 2009. Parallel Frequent Flier Program Partnerships: Impact on Frequency. Working Paper. University of California, Irvine.

Borenstein, S., 1989. Hubs and High Fares: Airport Dominance and Market Power in the U.S. Airline Industry *Rand Journal of Economics* 20, 344-365.

Borenstein, S. and N.L. Rose, 1994. Competition and Price Dispersion in the U.S. Airline Industry *Journal of Political Economy* 102, 653-683.

Cox, D., 1970. Analysis of Binary Data. London: Methuen.

Escobari, D., 2009. Systematic Peak-load Pricing, Congestion Premia and Demand Diverting: Empirical Evidence. *Economics Letters* 103, 59-61.

Gerardi, K.S. and A.H. Shapiro, 2009. Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry. *Journal of Political Economy* 117, 1-37.

Lederman, M., 2007. Do Enhacements to Loyalty Programs Affect Demand? The Impact of International Frequent Flyer Partnerships on Domestic Airline Demand. *Rand Journal of Economics* 38, 1134-1158.

Lederman, M., 2008. Are Frequent Flyer Programs a Cause of the 'Hub Premium'. *Journal of Economics and Management Strategy* 17, 35-66.

Lee, D. and M.J. Luego Prado, 2005. The Impact of Passenger Mix on Reported Hub Premiums in the U.S. Airline Industry. *Southern Economic Journal* 72, 372-394.

Mason, G. and N. Baker, 1996. Buy Now Fly Later: An Investigation of Airline Frequent Flyer Programs. *Tourism Management* 17, 219-232.

Table 1: Summary Statistics

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | mean | sd | min | max |
| Dependent Variable | | | | |
| *PROPFFT* | 0.0485 | 0.0655 | 0.00 | 0.934 |
| Fare Variables | | | | |
| *MEANFARE* | 180.9 | 65.76 | 11.05 | 774.9 |
| *20PCTFARE* | 113.7 | 41.55 | 0.450 | 676 |
| *80PCTFARE* | 238.4 | 113.1 | 1.450 | 1,178 |
| Characteristics / Controls | | | | |
| *MILES* | 1,288 | 780.1 | 67 | 6,089 |
| *DEPAHUB* | 0.204 | 0.403 | 0.00 | 1.00 |
| *PROPDEST* | 0.422 | 0.261 | 0.00532 | 1.00 |
| *PROPDEPA* | 0.183 | 0.195 | 1.67e-05 | 1.00 |
| *PROPDIRECT* | 0.453 | 0.457 | 0.00 | 1.00 |
| *LOADFACT* | 0.707 | 0.109 | 0.00 | 1.00 |
| Instruments | | | | |
| *CONNECT* | 0.637 | 0.481 | 0.00 | 1.00 |
| *NUMCONN* | 2.180 | 2.171 | 0.00 | 56 |

Notes: An observation is an airline $i$ in route $j$ during quarter $t$. The sample is from 1993:1 to 2009:3 and consists on 474,856 observations.

Table 2: Regression Results, Linear Model

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| VARIABLES | Pooled | RE | Within | Within |
| *MEANFARE* | 0.0245* | 0.0423* | 0.0418* | 0.0393* |
| | (0.000218) | (0.000222) | (0.000859) | (0.000854) |
| *MILES/10*$^3$ | 1.483* | 0.394* | | |
| | (0.0239) | (0.0459) | | |
| *DEPAHUB* | 0.303* | 0.637* | | |
| | (0.0269) | (0.0753) | | |
| *YEAR* | -0.394* | -0.453* | -0.350* | |
| | (0.00285) | (0.00216) | (0.00858) | |
| *PROPDEST* | 1.114* | 0.566* | 1.530* | 1.706* |
| | (0.0619) | (0.0647) | (0.278) | (0.276) |
| *PROPDEPA* | 0.989* | 0.973* | 1.756* | 0.557 |
| | (0.0923) | (0.0919) | (0.418) | (0.430) |
| *PROPDIRECT* | -1.756* | -1.780* | -2.566* | -2.614* |
| | (0.0271) | (0.0408) | (0.209) | (0.209) |
| *LOADFACT* | 0.150 | -1.518* | -0.895* | -1.444* |
| | (0.109) | (0.0949) | (0.241) | (0.254) |
| Observations | 474856 | 474856 | 474856 | 474856 |
| R-squared | 0.415 | 0.311 | 0.320 | 0.374 |
| Route/Carrier FE | No | No | Yes | Yes |
| Time FE | No | No | No | Yes |

Notes: The dependent variable is $F^{-1}(P_{ijt}) = P_{ijt} \times 100$, where $P_{ijt} = PROPFFT_{ijt}$. Figures in parentheses are robust standard error. ‡ significant at 10%; † significant at 5%; $^*$ significant at 1%. Using the total number of tickets in the route/carrier as weights.

Table 3: Regression Results, Log-odds Ratio

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | MLE | MLE | MLE |
| *MEANFARE* | 0.00203* | | |
| | (3.21e-06) | | |
| *20PCTFARE* | | 0.00290* | |
| | | (6.38e-06) | |
| *80PCTFARE* | | | 0.000675* |
| | | | (1.50e-06) |
| $MILES/10^3$ | 0.164* | 0.205* | 0.243* |
| | (0.000356) | (0.000365) | (0.000312) |
| *DEPAHUB* | 0.0794* | 0.103* | 0.0838* |
| | (0.000469) | (0.000463) | (0.000467) |
| *PROPDEST* | 0.245* | 0.268* | 0.261* |
| | (0.00137) | (0.00135) | (0.00136) |
| *PROPDEPA* | -0.0443* | -0.000770 | -0.00438* |
| | (0.00157) | (0.00154) | (0.00155) |
| *PROPDIRECT* | -0.186* | -0.137* | -0.175* |
| | (0.000588) | (0.000590) | (0.000592) |
| *LOADFACT* | 0.121* | -0.219* | -0.0296* |
| | (0.00240) | (0.00222) | (0.00235) |
| Observations | 474,856 | 474,856 | 474,856 |

Notes: The dependent variable is $F^{-1}(P_{ijt}) = \log\left(\frac{P_{ijt}}{1-P_{ijt}}\right)$, where $P_{ijt} = PROPFFT_{ijt}$. Figures in parentheses are robust standard error. ‡ significant at 10%; † significant at 5%; $^*$ significant at 1%. All specifications include time FE.

Table 4: Regression Results, Cox

| VARIABLES | (1) Pooled | (2) RE | (3) Within | (4) Within |
|---|---|---|---|---|
| *MEANFARE* | 0.00673* | 0.00753* | 0.00957* | 0.00834* |
| | (5.82e-05) | (3.81e-05) | (0.000236) | (0.000233) |
| *MILES/$10^3$* | 0.190* | 0.0883* | | |
| | (0.00614) | (0.00628) | | |
| *DEPAHUB* | 0.0490* | 0.0541* | | |
| | (0.00929) | (0.0131) | | |
| *YEAR* | -0.154* | -0.128* | -0.136* | |
| | (0.000932) | (0.000450) | (0.00281) | |
| *PROPDEST* | -0.0373 | -0.0737* | 0.534* | 0.573* |
| | (0.0228) | (0.0134) | (0.102) | (0.0990) |
| *PROPDEPA* | 0.571* | 0.564* | 1.296* | 0.872* |
| | (0.0261) | (0.0183) | (0.119) | (0.116) |
| *PROPDIRECT* | -0.796* | -0.675* | -0.818* | -0.847* |
| | (0.00692) | (0.00782) | (0.0432) | (0.0399) |
| *LOADFACT* | 0.103* | -0.824* | -0.775* | -0.787* |
| | (0.0367) | (0.0208) | (0.0793) | (0.0801) |
| Observations | 474856 | 474856 | 474856 | 474856 |
| R-squared | 0.528 | 0.453 | 0.400 | 0.504 |
| Route/Carrier FE | No | No | Yes | Yes |
| Time FE | No | No | No | Yes |

Notes: The dependent variable is $F^{-1}(P_{ijt}) = \log\left(\frac{P_{ijt}+a_{ijt}}{1-P_{ijt}+a_{ijt}}\right)$, where $P_{ijt} = PROPFFT_{ijt}$. Figures in parentheses are robust standard error. ‡ significant at 10%; † significant at 5%; $^*$ significant at 1%. Using the total number of tickets in the route/carrier as weights.

Table 5: Regression Results, 2SLS with Cox

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | 2SLS | 2SLS | 2SLS |
| *MEANFARE* | 0.0270* | | |
| | (0.00629) | | |
| *20PCTFARE* | | 0.0621* | |
| | | (0.0141) | |
| *80PCTFARE* | | | 0.0122* |
| | | | (0.00287) |
| *PROPDEST* | -0.00174 | 0.166 | 0.221 |
| | (0.216) | (0.180) | (0.172) |
| *PROPDEPA* | 0.529* | 0.0806 | 0.656* |
| | (0.176) | (0.256) | (0.157) |
| *PROPDIRECT* | -0.758* | -0.732* | -0.761* |
| | (0.0544) | (0.0576) | (0.0543) |
| *LOADFACT* | 1.554† | 0.330 | 1.344‡ |
| | (0.783) | (0.489) | (0.739) |
| Observations | 474856 | 474856 | 474856 |
| R-squared | 0.465 | 0.465 | 0.465 |

Notes: The dependent variable is $F^{-1}(P_{ijt}) = \log\left(\frac{P_{ijt}+a_{ijt}}{1-P_{ijt}+a_{ijt}}\right)$, where $P_{ijt} = PROPFFT_{ijt}$. Figures in parentheses are robust standard error. ‡ significant at 10%; † significant at 5%; $^*$ significant at 1%. Using the total number of tickets in the route/carrier as weights. All specifications include route/carrier FE and time FE.

Table 6: First Stage Regressions, 2SLS

| | (1) | (2) | (3) |
|---|---|---|---|
| VARIABLES | *MEANFARE* | *20PCTFARE* | *80PCTFARE* |
| *CONNECT* | -0.830‡ | 1.039† | -4.704* |
| | (0.503) | (0.445) | (1.074) |
| *NUMCONN* | -1.320* | -0.573* | -2.811* |
| | (0.287) | (0.114) | (0.606) |
| Observations | 474856 | 474856 | 474856 |
| F stat: $\alpha_1 = \alpha_2 = 0$ | 11.80 | 16.14 | 19.48 |
| p value: $\alpha_1 = \alpha_2 = 0$ | 7.52e-06 | 9.87e-08 | 3.51e-09 |

Notes: The dependent variable is shown below the column number. Figures in parentheses are robust standard error. ‡ significant at 10%; † significant at 5%; $^*$ significant at 1%. All specifications include route/carrier FE and time FE.