A close up of a piece of paper

Description automatically generated

MASTER THESIS

MSc Economics and Business

Specialization: Urban, Port, and Transport Economics

EXAMINING THE MODERATING EFFECT OF THE INTERNET ON U.S. AIRLINE FARES AND THE PREMIUM AFFORDED TO DOMINANT AIRLINES

EVAN HEITZMANN

467427

SUPERVISOR: YANNIS KERKEMEZOS

SECOND ASSESSOR:

DATE FINAL VERSION:

The views stated in this thesis are those of the author and not necessarily those of Erasmus School of Economics or Erasmus University Rotterdam

Table of Contents

[Section 1 – Introduction 1](#_Toc16861965)

[Section 2 – Literature Review 4](#_Toc16861966)

[2.1 Airport Dominance 4](#_Toc16861967)

[2.2 History and Controversy of Computer Reservation Systems 5](#_Toc16861968)

[2.3 Search Costs and Pricing 7](#_Toc16861969)

[2.4 Effect of the Internet 8](#_Toc16861970)

[Section 3 - Hypothesis Development 12](#_Toc16861971)

[3.1 Reconsidering the Hub Premium 12](#_Toc16861972)

[Section 4 – Data and Methodology 15](#_Toc16861973)

[4.1 Description of Data Sources 15](#_Toc16861974)

[4.2 Data Aggregation 16](#_Toc16861975)

[4.3 Pricing Equation and Summary Statistics 18](#_Toc16861976)

[4.4 Endogeneity 22](#_Toc16861977)

[4.5 Summary Statistics 24](#_Toc16861978)

[Section 5 – Results 27](#_Toc16861979)

[5.1 Models 1 and 2 27](#_Toc16861980)

[5.2 Model 3 29](#_Toc16861981)

[5.3 Extensions 32](#_Toc16861982)

[5.3.1 Extension 1 32](#_Toc16861983)

[5.3.2 Extension 2 33](#_Toc16861984)

[5.3.3 Extension 3 33](#_Toc16861985)

[Section 6 – Concluding Remarks and Discussion 34](#_Toc16861986)

[Appendix 1: Summary Statistics and Correlation Matrix 37](#_Toc16861987)

[Appendix 2: Miscellaneous Tables 38](#_Toc16861988)

[References 41](#_Toc16861989)

.

ABSTRACT

Study of the “hub premium” or instances of airport dominance that lead to super-competitive prices are prevalent in the airline economics literature. Early work theorized that it results from access (or exclusion) to airport facilities, frequent flyer programs, or bias from travel agents and computer reservation systems. Subsequent literature has focused on the first two factors and ignored the third. This thesis attempts to fill the gap in the literature by researching the effect of the internet as a moderating effect on the hub premium. Some evidence indicates that the pricing premium has declined since the 1990’s and 2000’s when most of the research was consolidated. We contend that decreased search costs and increased price and product transparency accruing from the internet could explain this decrease, however the results suggest this is not the case. A negative effect is found, but the results are unsubstantial, indicating that the internet has not played a significant role in decreasing the hub premium and that the other two factors in the literature are much more important market forces.

# Section 1 – Introduction

After the United States deregulated the airline market in 1978, many airlines reorganized their networks around hubs to minimize the number of flights necessary to serve the largest possible market. Among other changes, the following period was marked with an increase in firm mergers and consolidations. The consequences of these changes are still felt today, with implications for market power. Suppose a passenger wishes to book a flight in the US – there are a limited number of legacy carriers compared to the number of national carriers in Europe. And while mergers might lead to larger and more efficient networks, the effects of route dominance are more likely to outweigh these benefits. One of the motivating factors for deregulation was the theory of contestable markets, the idea that prices above competitive levels in monopolized routes simply could not exist – other carriers would enter the market to undercut the existing carrier’s prices. In reality, airline markets do not work this way and contestability has remained elusive (Borenstein, 1989). As a result, dominance in airline markets does lead to super-competitive prices.

The earliest work on the topic of concentration in the airline industry began with Borenstein (1989), who finds that both route dominance and airport dominance allow an airline to charge premiums to their passengers. Other researchers followed his example in examining what was termed the “hub premium” – fare prices at hub airports which were significantly higher than those in non-hub airports. Most research concerning the hub premium is conducted using data from the 1990’s or 2000’s and focuses on airport facilities (Ciliberto & Williams, 2010) or frequent flyer programs (Lederman, 2008), arguing that customers are locked-in with dominant airlines because they have the most extensive and convenient network and therefore are able to realize better rewards, or because lack of competition insulates market power and higher fares. Other researchers (Lee & Luengo-Prado, 2005) found that flights to certain cities attract a higher percentage of business passengers, providing an upward bias on previous estimates of the hub premium.

However, early work on the hub premium theorized bias from computer reservation systems and travel agents could exacerbate the problem. (Borenstein, 1989) To date, there has been no research exploring these potential factors which influence the hub premium, which this thesis seeks to remedy. Throughout the 1990’s passengers who wished to fly could either call each airline and request a price or seek a travel agent to sort through the potential itineraries to find a suitable product and fare. Superficially, it appeared that the customer received a good deal – he or she could trust the travel agent to work on their behalf to secure an appropriate fare. Except the truth of the matter was further from this ideal than the consumer imagined. Travel agents sifted through computer reservation systems designed and operated by airlines, and if these didn’t bias search results outright then a host of other factors lead to unfavorable outcomes. Namely, travel agent commission overrides rewarded travel agents for frequent and repeat bookings. A lack of information and high search costs resulted in passengers that relied on travel agents who did not always have their best interests in mind.

We posit that increased usage of the internet during the years 1997-2003, when most Americans gained access in their homes and at work, could decrease consumer search costs. Whereas the effect of search costs in previous years could only be theoretically modelled, these years provide the setting for a natural experiment in which, within a short time, many Americans gained the ability to more easily compare prices and products. Research into the effect of internet usage (Orlov, 2011) shows that increased usage rates can cause a direct decrease in airline prices. This thesis is intended to examine whether the hub premium is consistent over time or if it decreases due to the effect of the internet. The internet could allow passengers to avoid potential travel agent bias, and increased transparency of prices would allow them to more easily select the fares or products that suit them. We argue that in addition to the direct decrease of airfares due to the effect of the internet, there is an indirect effect of the internet reducing the hub premium. That is, not only could passengers choose lower fares within the dominant airline, they could more easily select the fares of non-dominant airlines. Because the hub premium results from an airline’s market dominance, this potential leakage of passengers to other airlines could result in a lower ability to charge a premium.

This thesis is the only research to directly consider the effect of the internet and reduced impact of computer reservation systems and travel agents on the hub premium. The data is a quarterly sample of 10% of all airfares in the United States and we focus our analysis in hub airports. We utilize an instrumental variables fixed effects model and find a significant and negative interaction effect between airport shares and a measure of internet usage. This implies that for a larger percentage share of departing passengers a carrier has of both endpoint airports in a market, there will be a more negative effect of the internet on fare prices. Conversely, as the rate of internet usage increases there is a decreasing effect of airport dominance. This implies that in hub airports a greater level of internet usage decreases the hub premium. However, the magnitude of the decreases is less than $2 for higher levels of internet usage compared to low levels of internet usage. Furthermore, we find that the interaction is only significant and negative at hub airports; there is no effect at non-hub airports, suggesting that the higher prices at hub airports incentivize passengers to search more thoroughly. These results imply that the internet has not had a sizable influence on the hub premium and that frequent flyer miles and access to airport facilities most likely play a greater role.

The remainder of the thesis is organized as follows: Section 2, an introduction to the relevant literature; Section 3, a formal outline of the hypotheses to be tested in this research; Section 4, a description of the data sources and summary statistics; Section 5, explanations of the main results; and Section 6, concluding remarks and a final discussion.

# Section 2 – Literature Review

The purpose of this section is to illuminate the research previously conducted on (2.1) airport dominance and sources of market power that contribute to airline ability to charge a hub premium, (2.2) the history of computer reservation systems and how they presented passengers with problems of asymmetric information, (2.3) a formal explanation of search cost theory, and (2.4) an assessment of the literature regarding the effects of the internet on airline prices and price dispersion.

## 2.1 Airport Dominance

The deregulation of the U.S. airline industry resulted in numerous changes concerning the prominent business model of most carriers. Whereas new low-cost carriers competed with point to point networks, legacy carriers reorganized their networks around airport hubs. There is strong evidence that hub and spoke models increase economies of density along spoke routes, so the average cost per passenger should fall because airlines are able to more easily fill their planes. (Brueckner, Dyer, & Spiller, 1992) This, in turn, should lead to lower prices. However, routes which either originate or have their destination in hub airports remain outliers to this effect. There is strong statistical evidence to suggest that airport dominance leads to market power for an airline. In the literature, this effect is sometimes referred to as the “hub premium” – the ability of the dominant airline to charge higher prices for hub-originating or hub-terminating flights. How can this happen? Borenstein (1989) suggests frequent flyer programs, travel agent commission overrides, and improper use of computer reservation systems. Of these three things, the effects of commission overrides and reservation systems are most likely significantly reduced in magnitude due to changing consumer demand and structural changes in ticket distribution systems.

Subsequent research by Evans and Kessides (1993) shows that route dominance does not contribute as strongly towards the hub premium as airport dominance. In fact, the effect of route dominance is not statistically different from zero. Lederman (2008) finds that frequent flyer programs constitute a significant portion (between 25% and 37%) of the hub premium. This implies passengers (especially business class passengers) are willing to pay more for a utility amenity or benefit (e.g., increased connectivity, more direct flights, more frequency). Passengers have an incentive to choose the airline with the most extensive network, as frequent flyer programs reward passengers who purchase flights with one carrier. Ciliberto and Williams (2010) add to the literature by arguing that access to gates, sublease fees, and majority-in-interest clauses can limit entry of potential competitors, forming an additional advantage for the hub-operating airline. Snider and Williams (2015) confirm this by analyzing the passage of the Wendell H. Ford Aviation Investment and Reform Act for the 21st Century (AIR-21) to show that increased access to facilities results in a decrease in average fares. AIR-21 stipulated that highly concentrated airports (where two airlines retain more than 50% of airport departures) must “take concrete steps to ensure that new entrants had ample access to airport facilities.” (Snider and Williams, p. 1002) Using a regression discontinuity design, they find that airports just above the 50% threshold experience a decrease in prices.

Exactly how high are these hub premiums? Borenstein (1989) uses an instrumental variables approach without fixed effects and finds that a one percent increase in his measure of airport dominance increases the median fare by 0.12%. Evans and Kessides (1993) state clearly that once introducing fixed effects, their model shows a one standard deviation increase in airport dominance results in an average fare increase of about $18. Lederman (2008) utilizes a dummy variable approach and finds that on average, flights that depart or arrive in an airline’s hub are priced between 7% and 18% higher than the fares charged by other carriers.

Robust evidence supports the claims that access to facilities and airline rewards programs contribute to the fare premium charged by dominant airlines, and yet Borenstein (1989) writes that travel agent and computer system bias “may have as large or larger effects.” (p. 346).

## 2.2 History and Controversy of Computer Reservation Systems

In the early 1990’s tickets were almost entirely purchased through travel agents; the destabilizing force of the internet had not yet arrived to provide alternative ticket distribution channels. By 1999 travel agents still dominated most of the market – Ravich (2004) notes that 75% of tickets were sold through agents. Of these tickets, more that 90% were distributed through computer reservation systems (CRS) that the airlines provided. Concerns were voiced regarding the potential for the system’s algorithm to bias search results in favor of the airline’s own tickets. Other than direct bias of the list, the reservation system could deceive the customer by simply failing to list competing tickets (Minick 2000). Regulation from the Department of Transportation specifically targeted this type of anti-competitive behavior, which was rampant in the 1980’s. Although some still claimed that airlines could discreetly engage in this type of behavior, it’s presence remains unlikely through 1990’s.

There are, however, additional unregulated ways in which a CRS could bias ticket purchases. Travel commission override programs are the most prominently mentioned in the literature – they are essentially rewards programs for travel agents. Agents who book large volumes of flights with airlines that own the CRS have the potential to earn bonuses. Minick (2000) details three further biasing tools the CRS owning airline utilized: travel agents could gain access to special lounges and VIP rooms and services in airports, they could have the power to overbook last minute customers and earn greater commissions from these transactions, and they could be awarded free flights or special promotions and discounts from the airline.[[1]](#footnote-1) As Minick (2000) succinctly concludes, “Many people see travel agents as their agents, not the agents of the CRS. The influence of the CRS inhibits a travel agent’s ability to provide objective advice.” (p. 904)

Furthermore, ownership and participation in reservation systems could limit consumer welfare. Airlines recognized that owning and distributing a reservation system can be beneficial, and yet developing a sophisticated computer program and marketing it to travel agents is not an insignificant cost. Small carriers remained at a disadvantage and required listing their fares on the reservation systems of larger airlines. Minick (2000) notes that the co-host agreements that large CRS owning airlines imposed on smaller airlines were imbalanced with higher booking fees and surcharges. The additional fees were even higher when the small carrier was a competitor on the same route. The higher costs for these airlines carried over into the total ticket price for the passenger. In effect, the true fares of low-cost carriers and small airlines was being artificially distorted by reservation systems and therefore passengers motivated by price might have less incentive to choose their tickets. The hub premium can be distilled merely as the price effects of factors which increase an airline’s share of passengers in an airport. Because the co-host agreements could potentially lead to more passengers choosing tickets of the larger airline (due to decreased price dispersion), it is reasonable to assume that the subsequent insulation from passenger leakage could influence the hub premium. As a result of this uneven playing field and travel agent bias, computer reservation systems led to anti-competitive markets and disadvantaged consumers.

Regulation concerning computer reservation systems ended in 2004, not because of any change from within the systems, but because the internet had rendered them fairly useless. By this time, airlines could build their own websites and market tickets directly to passengers, avoiding the costly commission fees that travel agents charged. Ravich (2004) writes that the addition of the internet cut costs of distributing tickets by about half, “often eliminating travel agent services, lowering transaction costs, and allowing airlines to fill otherwise empty seats through low-price Internet deals.” (p. 400) The previous quotation should not imply that all travel agent services were eliminated. In fact, at the turn of the century they still accounted for around half of all ticket sales because businesses and passengers relied on their expertise for booking complicated itineraries. (Orlov, 2011) Around this time, airlines started simultaneously decreasing the amount of commissions they paid to agents, a move which was marketed as leading to lower prices for consumers. (In 2002, travel agents switched business models and began charging their commissions to passengers.) The decline in sales was led in part because travel agents utilized global distributions systems (GDSs) to access fares. Fares published on airline websites, however, were both lower in price and inaccessible to travel agents. (Orlov, 2011)

## 2.3 Search Costs and Pricing

In many fields of economics, information or access to it plays a key role in market behavior.

One of the requirements for a textbook example of a perfectly competitive market is that buyers have perfect information about products or can acquire this information without cost. Intuitively we know this to be true, shoppers in grocery stores will naturally move to where cashier lines are shortest. However, passengers looking to purchase airline tickets do not face zero search costs, hence the existence of the entire travel agent industry. If a person looking to buy a ticket did not go through an agent, he or she would have to call or visit the ticket counter of every airline and request a ticket price. Stahl (1989) provides a formal theoretical outline of search costs and buyer outcomes in an oligopolistic market with *N* firms. Buyers compare the prices of each firm, and those with zero search costs can endlessly search and arrive at a competitive equilibrium. However, passengers with positive search costs arrive at monopolistic prices. This occurs because buyers eventually settle on prices that match pre-determined reservation values. Theoretical explanations of consumer search costs (such as Stahl (1989)) anticipate that while the effect on price dispersion is ambiguous, the effect of reduced search costs will undoubtedly shift average prices lower.[[2]](#footnote-2)

The introduction of the internet plausibly lowers search costs, and according to this model induces buyers in the market to arrive at more competitive prices. Ultimately this is an assumption of the author, but it is by no means a weak or unfounded assumption. In the past, the effect of decreasing search costs could only be theoretically explained. Differences in internet usage present a unique opportunity to empirically explore whether online consumers are in fact more informed and pay lower prices. Brown and Goolsbee (2002) compare life insurance premiums in areas that have different internet usage rates and find that premiums are lower when there is high internet penetration. Importantly, only the life insurance plans that were available online experienced decreases in prices, and these decreases only occurred during the years when the plans were available for comparison online. It is only reasonable to extend this line of analysis and apply it to airline prices. Airline passengers, who in the past mostly relied on the advice of travel agents, could now more easily compare prices online for themselves. The following section summarizes the research into the internet’s effect on airfares.

## 2.4 Effect of the Internet

Berry and Jia (2008) write that in 1996 only 0.5% of tickets were sold online. 2007 online sales reached “as high as 50-60% in US”. The percentage has most likely only increased since 2007 as reputations of low-cost carriers improve, the sizes of their networks expand, and passengers feel more comfortable (and secure) booking on the internet. Most passengers no longer book with traditional travel agents, reducing consumers’ exposure to bias due to commission overrides and previously mentioned agent disincentives. The internet’s ability to reduce search costs (through third party booking services like Expedia) has made passengers more price sensitive. Despite greater demand for air travel in 2006 than 1999, Berry and Jia (2008) find that consumers shifted preferences toward low- and medium-priced tickets and away from high priced products. Furthermore, Dana and Orlov (2009) show strong statistical evidence that internet penetration rates are correlated with airlines’ ability to increase their load factors. This suggests that consumers can more easily find the itineraries or products that suit their demand, increasing allocative efficiency.

Price effects of the internet have only been directly considered in a few papers and with mixed results. Orlov (2011) uses data on household internet adoption rates and finds that an increase in internet usage reduces average prices and increases intrafirm price dispersion. Even when controlling for route characteristics and the presence of low-cost carriers, Orlov (2011) shows that the internet decreases prices between 3 and 5 percent, and that the internet does *not* significantly affect price dispersion between firms. Lane and Verlinda (2004) find conflicting evidence that increased internet usage does in fact significantly affect interfirm price dispersion. They conclude that the internet increases the spread between prices of restricted fares and unrestricted fares. Contrary to the majority of the literature, they find that the internet has an uncertain effect on unrestricted fares, but that it *increases* restricted fares. This is argued as the effect of product differentiation and the result of firms competing over quality for the highest spending passengers. Evidence from Granados, Gupta, and Kauffman (2012) support this claim by finding that whereas decreased search costs relating to price will increase passenger price elasticity, increased product differentiation leads to more price inelastic passengers. Regarding the increase in restricted fares, Lane and Verlinda (2004) even admit skepticism regarding this finding, and state that it may be the result of “low precision.”

Brunger (2010), however, utilizes detailed data on individual transactions for Continental Airlines and finds that tickets for passengers who are “clearly leisure” are significantly affected by the internet. Even on routes where the same fares are available to both online shoppers and brick-and-mortar travel agents, passengers who book online choose for fares which are 3 to 8 percent lower than the passengers who book through a travel agent. Despite being a case study and acknowledging the potential lack of applicability to other airlines, Brunger’s (2010) findings are corroborated by both Orlov (2011) and Sengupta and Williams (2014). The paper by Sengupta and Williams (2014) finds that “controlling for ticket characteristics, capacity, route, and carrier fixed effects, online customers on average pay roughly 11 percent less than offline customers.” Furthermore, the internet effect is solely confined to the tickets purchased online, i.e., only passengers who book online find lower prices. The prices of other shoppers are not affected by the presence of the internet. Interestingly, their paper also considers the effects of market structure on airline prices by including variables for market share and a dummy variable to indicate if an airport is a hub. The results show that these measures of airport dominance have no statistically significant effect on airfares. This contrasts with Brunger (2010) who also includes a dummy variable for a hub airport and finds the effect of the hub increasing prices by about four percent.

Can we expect that the internet effect is confined only to a certain type of traveler? Brunger (2010) only analyzes leisure passengers, and Sengupta and Williams (2014) do not account for any difference in leisure or business characteristics. Castillo-Manzano and Lopez-Valpuesta (2010) utilize 2006 survey data from four Spanish airports to isolate variables that cause a respondent to be more likely to book a ticket online. They write that “according to the results that were obtained, the profile of passengers who are more likely to make their bookings online is that of a young person (of between 15 and 30 years of age), more likely to be female, a student or with a high academic level, a habitual traveler, who is booking a trip that is not very complex or is to a destination that is already known and, in the main, a user of LCCs.” Alternatively, older passengers and those who are making a short business trip were more likely to book their ticket directly with the airline or through a traditional travel agent. Business passengers can be expected to exhibit more price insensitive behaviors than leisure travelers. On one hand, airlines have known for a long time that business travelers are more likely to book flights shortly before departure dates and that leisure travelers book further in advance, indicating that leisure travelers know their plans in advance and can afford to spend time to compare flights. Additionally, leisure passengers most probably bear the full weight of their travel costs. A business traveler may be motivated by the accrual of frequent flier points and, because the employer likely purchases the ticket, is not as incentivized to find the lowest cost flight, instead prioritizing departure times or arrival times.

A seminal paper by Mason (2000) explored the utility benefits for business passengers of using low-cost carrier, inspiring further work by Fourie and Lubbe (2006), Huse and Evangelho (2007), and Evangelho (2007). Mason (2000) uses survey data at British airports and finds that contrary to expectations, ticket price and not rewards points or departure times are the prominent motivating purchase factor for business passengers. The expansion of low-cost carriers has certainly tapped into latent travel demand for leisure passengers, but Mason (2000) argues that low-cost carriers also have stolen passengers from legacy carriers and incentivized worker travel for small and mid-size firms. Indeed, Mason (2000) writes that “easyJet have indicated that on some routes the proportion of business traffic achieved is as high as 50%” (p. 109). Both Mason (2000) and Alamdari (2002) claim that business passenger traffic on low-cost carriers may not be too surprising as companies adopt travel policies to restrain employee behavior. Alamdari (2002) notes that survey data shows “some 79% of companies have a travel policy” (p. 342) While the policy alone may not curb inefficient purchasing behavior (see Douglas and Lubbe (2008) for factors that may lead to policy violation), it indicates that the majority of employees who travel are subject to some form of purchasing control. Business travelers, therefore, may still benefit from the increase in price transparency offered by websites.

# Section 3 - Hypothesis Development

While the previous section was dedicated to reviewing literature concerning computer reservation systems, travel agents, and the moderating effect of the internet, this section will motivate revisiting the effect of the hub premium and formally state the hypotheses to be tested in this thesis.

## 3.1 Reconsidering the Hub Premium

A working paper by Borenstein (2005) compares the average fares at dominated airports and competitive airports and finds that airport premiums have been declining over the time period 1995-2004. The calculation involved is rather undeveloped, Borenstein simply takes the average airfare for an airport along each route. The airfare is classified according to which distance category it belongs to, e.g., 401-450 miles. The premium, therefore, is the percent paid in excess of the national average for the same distance category. The results, however, are rather compelling and show that “after adjusting for inflation (using the consumer price index, all urban) the [average ticket price] decline has been quite dramatic, more than 20%.” (p. 2) Additionally, Borenstein (2005) finds that the dispersion in prices between the most expensive airports and least expensive airports decreased between the same period – the hub premiums at most airports declined over the years in sample, but some of the cheapest airports rose in average price as well. As the literature shows, the decrease in fares is most likely attributed to the market entrance and expansion of low-cost carriers. (Brueckner, Lee, & Singer (2013) accurately describes the effect of low-cost carriers on airport prices.) Alternatively, Berry and Jia (2008) explain that from 1999 to 2007, airlines were able to increase their load factors by about 10%. A plane that has sold more seats is able to spread the average costs more evenly between each passenger and potentially enable the airline to lower ticket prices. These two forces are probable explanations for the decrease in fares, yet other factors possibly exist and have not been reconciled with previous literature.

Most empirical research on the hub premium attempts to estimate a reduced form price equation, where ticket fares are the dependent variable which are explained by ticket and market control variables and a measure of airport dominance either identified as 1) airport market shares or 2) a dummy variable indicating a hub airport. The main purpose of this thesis is to posit that growing internet adoption rates in the beginning of the century constitutes an omitted variable, biasing previous research. For omitted variable bias to exist, there must be a confounding effect in the error term that both correlates with the dependent variable (price) and the explanatory variable (airport dominance). Previous empirical literature from Section 2.4 readily shows that growth rates in internet usage and online purchasing behavior is correlated with price. Correlation between internet usage and airport dominance is not so easily proven, primarily because it has never been directly examined. The literature in Section 2.4 shows conflicting evidence regarding the effect of the internet on interfirm price dispersion. However, Evans and Kessides (1993) include a variable which measures an airline’s share of tickets sold on their own computer reservation system (i.e., measuring the effect of CRS ownership and any effect on market power.) Inclusion of the variable lowers the coefficient associated with airport dominance, which demonstrates a correlation between reservation system ownership and price. As we expect that the effect of CRS ownership has been eliminated and passengers can more easily compare prices, the effect of airport dominance should in turn decrease. Brunger (2010) lends support to this, as the model in their paper finds the coefficient for online purchases changes with the inclusion of a dummy variable indicating hubness, indicating a correlation between internet usage and airport dominance. Furthermore, as low-cost carriers directly market their products to passengers through self-owned websites, increased travel depends on internet access. Escobar-Rodrígueza and Carvajal-Trujillo (2014) write that not only are low costs an incentive for LCC passengers, but also the overall ease of online booking attracts customers. To state it in plain language, we expect a negative correlation between internet usage and airport dominance due to reduced search costs and increased ease of booking. As Sections 2.2 and 2.3 explained, positive search costs prevent consumers from arriving at competitive prices because they purchase self-determined acceptable prices. Prior to the introduction of the internet consumers relied on travel agents to complete their bookings. From the consumer’s perspective, doing business with a travel agent mitigates the extra costs associated with imperfect information. However, due to travel agent and computer reservation system bias, there is reason to believe that these bookings do not represent optimal prices for consumers. The internet, in turn, could facilitate greater information and comparison of ticket prices between airlines. A consumer, therefore, could be more likely to purchase the ticket of a low-cost carrier or even the fare of a non-dominant legacy airline in a hub. Because the hub premium in the literature is the result of high airport market shares, the decrease in market share for a dominant airline due to the internet should result in a decrease in the pricing premium. Thus, we formulate the main hypothesis of this research:

H1: Household internet usage in metropolitan areas decreases the pricing premium associated with airport dominance in hub airports.

This hypothesis presupposes that omitted variable bias in previous literature has led to an upward bias in the coefficient reported for airport dominance. For this to be the case, we expect the following to also be true:

H2: An increase in the percentage of airport market shares leads to an increase in the itinerary fare.

H3: An increase in the percentage of households who access the internet leads to a decrease in itinerary fare.

# Section 4 – Data and Methodology

This section will describe the databases accessed for this research, how the data is consolidated, identify key variables included in the study along with descriptive statistics, resolve any issues concerning endogeneity, and rationalize a model to test the hypothesis.

## 4.1 Description of Data Sources

This research, like many previous papers on the subject, analyzes the market for US domestic air travel. While data on European and other international markets exists, the US domestic market is chosen for ease of data accessibility and relevance to prior literature.

Data regarding prices and route characteristics is accessed from the Bureau of Transportation Statistics (BTS). For this research, we utilize the Airline Origin and Destination Survey (DB1B). As the BTS states, it is “a 10% sample of airline tickets from reporting carriers collected by the Office of Airline Information of the Bureau of Transportation Statistics. Data includes origin, destination and other itinerary details of passengers transported.”[[3]](#footnote-3) Data is available for every quarter from 1993 to 2018, however this research only aggregates years in which internet usage is expected to change most significantly: 1997-2003. The survey contains three individual datasets: DB1BCoupon, DB1BTicket, and DB1BMarket. Together, the master file contains information regarding origin airport, destination airport, fare class, ticketing carrier, and total airfare. In this research, a market is defined as each unique combination of origin and destination market endpoint airports.

Additionally, we utilize information from the T-100 Domestic Segment (All Carriers) database which is also collected by the Bureau of Transport Statistics.[[4]](#footnote-4) This contains monthly data regarding market traffic and available seat miles, which is used to calculate the load factors and number of passengers in a market.

Additional data regarding household internet adoption rates is available from the Computer and Internet Use supplement to the Current Population Survey (CPS), which has been accessed from IPUMS-CPS. The CPS is a monthly survey collected jointly by the United States Census Bureau and the Bureau of Labor Statistics. IPUMS-CPS aggregates and harmonizes the variables and coding between survey waves and publishes CPS data in formats that are more accessible than those published by the Census Bureau. The supplement concerning Computer and Internet Use was collected during 1997, 1998, 2000, 2001, 2003, and 2007. Following from Orlov (2011), internet adoption rates are measured by respondent answers to a question of whether or not a person accesses the internet from home. The ratio of positive answers to the total number of survey respondents (including both “No” and “Not in Universe”) for a Metropolitan Statistical Area indicates the rate of internet penetration in the area. The IPUMS-CPS data is then collapsed by a variable which signifies the metropolitan area per year.

Lastly, this research interpolates information regarding Metropolitan Statistical Area (MSA) population (as designated by the United States Census Bureau.[[5]](#footnote-5)) The purpose of this information is to generate appropriate instrumental variables in the model.

## 4.2 Data Aggregation

To construct the final sample, we first collect each quarter of the DB1B for every subsample in the years 1997-2001 and 2003. To account for possible loss of internal validity resulting from decreased passenger traffic and industry-wide shocks due to the September 11th terrorist attacks, the entire year of 2002 and the last two quarters of 2001 are not included in the final sample. Certain steps were taken in each dataset to ensure there are no errors. In the DB1B Coupon and DB1B Market, observations are dropped if the origin airport is equal to the destination airport. Furthermore, observations in the DB1B\_Coupon are removed if the ticketing carrier is not equal to the operating carrier, in accordance with Gerardi and Shapiro (2009). Interlining flights and flights for which there is a codeshare agreement are not desired in the final sample. The DB1B coupon dataset contains information regarding fare class, and in accordance with steps taken in previous literature fare classes of non low-cost carriers are dropped if they are not economy class.[[6]](#footnote-6) Lastly, any observation is dropped if the flight distance is less than 5 miles. In the DB1B Ticket dataset, any observation is dropped if the itinerary fare is recorded as less than five dollars, the ticket is purchased as a bulk fare, or the dollar credibility (a binary variable which indicates the Department of Transportation’s confidence in the reported fare) is zero. In the DB1B Market, observations are dropped if the ticket is purchased as bulk fare, the market fare is below 15 dollars, there is a ticketing carrier change, the ticketing carrier is not equal to the operating carrier, the origin airport is equal to the destination airport, or the number of coupons in the market is above 2. This last requirement is to ensure that the itineraries under consideration at most only include those for which there is a flight from an origin, a connection in a hub airport, and a flight to the destination. Simultaneously, the T-100 is also prepared for merging. Observations in the T-100 segment database are dropped if the departures performed per route are equal to zero, the distance is less than five miles, or the seats available on the route are zero. (The recorded number of passengers can equal zero, as it conceivable that a flight with a small carrier does not sell tickets, but a flight with zero available seats does not sound reasonable, unless it is a small private plane.)

Population data in each MSA on a yearly basis is now only available on the Census Bureau website for the years 2010 through 2018. However, the 1990 census and 2000 census contain information about each MSA’s population. Values for the years 1997 to 1999 are linearly interpolated based on the known rate of change for each MSA, and the years 2001-2003 are extrapolated based on the same rate of change. The MSA population is then manually matched with the associated origin airport ID from the DB1B and the calculated internet adoption rate. Because the rate of internet usage is known only for 1997, 1998, 2000, 2001, 2003; the rates for 1999 and 2002 are taken as the average of the preceding and successive years. This is then merged with the T-100 Segment data on the basis of origin airport and year. A copy of the data is then merged to generate population data for the destination airport. The merged data is collapsed by an indicator variable which signifies the route, year, quarter, and reporting carrier code. Each observation, therefore, contains the average number of available seats, passengers, distance traveled, as well as the population at the endpoints of each airport. The collapsed T-100 Segment data is then merged with the DB1B Coupon so that each leg of a market coupon has an associated load factor.

Separately, the DB1B Ticket and DB1B Market datasets are merged to insert the total itinerary fare in the Market dataset. Duplicates in the itinerary ID are dropped to ensure that each itinerary fare is not counted twice and that each observation therefore contains information referring to the first market trip and not the return trip. The price of a flight from, for example, New York to New Orleans is not the same as a flight from New Orleans to New York. Information about origin population, destination population, and internet usage rates are then merged to the combined DB1B Market and Ticket samples. Finally, this is merged with the DB1B Coupon and T-100 file, and then collapsed by an indicator value so that each observation records the average fare (as well as the 20th percentile fare and the 80th percentile fare) and associated data along one route for one carrier in one year quarter combination.

Because this thesis is interested in the impact of airline dominance in hub airports, the final model is restricted only to hub airports.[[7]](#footnote-7) Borenstein (1989) and Evans and Kessides (1991) simply estimate the impact of increased airport dominance in all airports, but we would argue that this estimated coefficient is contaminated with the effect of non-hub airports. If there is strong indication from previous literature that hub airports charge a fare premium, then there is reason to believe that the dominance premium is higher in these airports. Furthermore, as per H1 this thesis is only intending to estimate the moderating effect of the internet on the dominance premium in hub airports.

## 4.3 Pricing Equation and Summary Statistics

The sample contains 140,518 observations spread across more than 40 carriers in 6 years. We first estimate the following fixed effects equation:

Equation 1:

for:

: 1997, 1998, 1999, 2000, 2001, 2003

In a decision that is consistent with the literature, the dependent variable for this research is the natural log of the average fare of airline *i* in route *k* at quarter *j.* The coefficient on *AirportShare* measures the average of carrier *i*’s percent of total departing passengers at both endpoints of market *k* in quarter *j*. It is measured on an interval of [0,100] to facilitate easier interpretation of the results. is a set of control variables, is the set of fixed effects for carrier *i* in route *k*, is a set of year fixed effects, and is the random error.

The second equation represents the general fixed effects model which is compared to Equation 1. If greater usage of the internet leads to a decreased premium, then we should expect that is negative and the combined effect of and leads to lower fares than those estimated with Equation 1.

Equation 2:

for:

: 1997, 1998, 1999, 2000, 2001, 2003

Equation 2 augments Equation 1 with a few additional variables of interest. The coefficient on *InternetRate* estimates the effect of a one percentage point increase or decrease in the ratio of households who use the internet in their home in the MSA of the origin airport of market *k* in quarter *j. InternetRate* is measured on an interval of [0,100]. The variable *Rate\*Share* is an interaction term between *InternetRate* and *AirportShare*, and is an interaction term between *InternetRate* and each year of the sample, which conditions for the effect of increased internet usage across time and in different years.

Study of the hub premium primarily originates with Borenstein (1989), but it is Evans and Kessides (1993) who introduced what is now standard practice of carrier, route, and time fixed effects which have continued to be used in most subsequent research. The purpose of a fixed effects model as opposed to OLS is that there are most likely variations in price that are unique to each route, and the fixed effects model controls for these time-invariant factors that are unique to all carriers. Carrier fixed effects are introduced to condition for the possibility that certain firms may adopt different pricing strategies which are carrier-unique and do not vary within the sample. Similarly, the use of year fixed effects controls for shocks to the entire airline industry which are not carrier unique, such as increases or decreases in the price of oil. Furthermore, the use of route fixed effects is important to control for the effect of flying into or out of certain airports. This potentially conditions for the effect of airports which more frequently experience delays, the effect of having an adjacent airport in the MSA, and the effect of the destination having a greater mix of business or leisure travelers. As mentioned, previous literature estimates the effect of airport dominance either with a dummy variable to indicate a hub airport or an interval measure of market share. The use of fixed effects precludes the presence of a hub dummy variable as it would be eliminated with other time invariant variation. A correlated random effects model, or a similar hybrid model, would allow for the inclusion of time-invariant effects into what is essentially a fixed effects model. The estimated beta coefficient on time-variant explanatory variables will be identical to the estimated beta under a fixed effects model, however the estimated beta coefficient for time-invariant variables would be similar in direction and magnitude to those estimated with a standard random effects model. Previous literature indicates this may not be ideal; a Hausman test in Evans and Kessides (1993) rejects the random effects model in favor of fixed effects, as the unexplained factors are most likely correlated with the set of explanatory variables. Therefore, a fixed effects model is preferred.

The set of control variables has its foundation in the same control variables used in Borenstein (1989) and Evans and Kessides (1993) as well as later literature that illustrates the significant effect of low-cost carriers. The variables are limited to:

* *LnDistance –* The natural log of the distance traveled by carrier *i* between the two airport endpoints of route *k* in quarter *j.* Similar to Borenstein (1989), the coefficient is expected to be positive, as greater distance implies greater fuel costs which would increase price.
* *LCCRoutes –* The number of routes in the origin airport of route *k* that are served by low cost carriers. A dummy variable indicates if the route is served by an LCC, which are defined as JetBlue Airways, ExpressJet Airlines, Frontier Airlines, AirTran Airways, Allegiant Air, Valujet Airlines, Spirit Airlines, Sun Country Airlines, and Southwest Airlines. The sign of the coefficient is expected to be negative, as increased competition from low cost carriers will tend to reduce the average fare in an airport. The effect of low-cost carriers is well reported in the literature, and legacy carriers significantly reduce fares in response to competition from them.
* *LoadFactor[[8]](#footnote-8)* – The average load factor in market *k* belonging to carrier *i* in quarter *j*.The coefficient is expected to be negative because flights which fill more seats are able to split the total costs over more passengers and potentially reduce ticket prices. Evans and Kessides (1993) presumably use the number of passengers in a route to proxy for this effect, but this research assumes that the actual load factor is a better measure, as it also indirectly controls for the effect of the size of the aircraft. Borenstein (1989) theorizes that the sign of the coefficient could also be positive as passengers recognize the inconvenience of full flights and so the airline compensates them for this burden. This does not appear to be a reasonable assumption, as passengers (or even the airline) cannot predict or know the load factor before the flight departs. The interval is from [0,100].
* *OriginPop* – The annual population of the MSA (scaled by 10,000 people) to which the origin airport in route *k* belongs. It is expected that a higher population will increase the demand for air travel products. Airlines will be able to exploit this by utilizing larger and more efficient aircraft, reducing the overall cost of passengers in the origin city. The expected sign is therefore negative.
* *DirectShare* – The percentage share of carrier *i* flights in route *k* of quarter *j* that are direct flights. The interval is [0,100]
* *RouteShare* – The percentage market share of carrier *i* in route *k* of quarter *j*. This is a common control variable included in the literature, and we suspect that fares can increase as a carrier captures a higher percentage of the flights in a route. The interval is [0,100].

An interaction term between *OriginPop* and *IntRate* (which could measure the probable effect that MSAs with higher populations could be more likely to have better digital infrastructure and therefore increased internet usage) was additionally considered, but not only did a preliminary output from the regression show highly insignificant results for the term, there was no correlation between this proposed term and *LnFare.* Another common measure included in previous literature is the share of LCC products within one route. While this is certainly negatively correlated with route fare, the alternate measure of *LCCroutes* is chosen because of correlation with *AirportShare*. Inclusion of both measures of LCC competition would most likely impede interpretation of the results as the variables could be highly correlated. In other words, *LCCroutes* already contains information about the presence of low-cost carriers on routes.

## 4.4 Endogeneity

Previous literature, and common sense, anticipates the influence of endogeneity on several variables in the right-hand side of the price equation. This would necessarily present an issue regarding the interpretation of results and the confidence with which the explanatory variables can be said to be causal. For the estimated coefficients on the explanatory variables to be unbiased, a necessary condition is that they are not correlated with variation in the error term. This research assumes that this condition has been met, and there are no remaining variables to include in the price equation which influence itinerary fare and are correlated with the independent variables. A second requirement for unbiased estimators is that there is no simultaneity bias. This has been readily accepted and accounted for in past literature concerning airport dominance, with a notable exception being Brueckner, Lee, and Singer (2013). The results of this research show that a Hausman test of endogenous regressors rejects the idea that certain variables can be treated as exogenous. The variables *AirportShare*, *LCCRoutes, Rate\*Share* and *LoadFactor* are assumed to face simultaneity bias; in the cases of the first two it is possible that a high average itinerary fare can draw entry into a route, especially by a low-cost carrier which is able to exploit the large difference in price to attract passengers. *Rate\*Share* is endogenous for the reason that it is composed of *AirportShare.* In the case of *LoadFactor,* lower itinerary prices potentially entice passengers to fly on a route. The variable *IntRate* is, in accordance with Orlov (2011), assumed not to face simultaneity bias. This is a reasonable assumption, as increases or decreases in airline prices most likely do not induce households to acquire internet access. The variable is additionally assumed to be exogenous to variation in the error term.

Preliminary results suggest that the IV estimation performs too weakly when all control variables are correctly instrumented. This thesis has thus correctly instrumented only for the main explanatory variables: *AirportShare* and *Rate\*Share*.[[9]](#footnote-9) These endogenous regressors are instrumented for according to a combination of variables from Borenstein (1989) and Gerardi and Shapiro (2009):

* *AvgEndPop –* The simple arithmetic mean of the populations of both MSA’s of the airports at endpoints of route *k* in quarter *j.* Population data is constructed on a yearly basis, so all four quarters within a year register the same value.
* *GeoEndPop –* The geometric mean of the populations of both MSA’s of the airports at endpoints of route *k* in quarter *j.*
* *LnPassenger* – The natural log of the number of enplaned passengers on route *k* in quarter *j*.
* *GENP* – An instrument similar in construction to the general enplanement instrument introduced in Borenstein (1989). It is calculated as: where the numerator represents the natural log of the total enplanements of carrier *i*  in the origin airport of route *k* in quarter *j*, and the denominator represents the natural log of the total enplanements of all other carriers (grouped as *u*) in the origin airport of route *k* in quarter *j*.
* *Avg2­* – The square of *AvgEndPop.*
* *Geo2* – The square of *GeoEndPop*
* Interaction terms between (1) *AvgEndPop* and *IntRate* and (2) *GENP* and *IntRate* to instrument for the term *Rate\*Share* which is by construction endogenous.

## 

## 4.5 Summary Statistics

A preliminary glance at the data reveals the following information in Table 1. The classification of hub airports follows from Lee and Luengo-Prado (2005). The table is organized by the final column in descending order. It shows that our data is concurrent with the literature and that the relative share of originating passengers in an airport matches those of Lee and Luengo-Prado (2005).

Table 1: Summary Statistics of Hub Airports



In all hub airports, the mean itinerary fare and share of passengers exceeds those of non-hubbing airports. This implies the existence of a hub premium, and yet further analysis is still required. Correlation between passenger share and fare price could simply be spurious, conditioning for market effects is necessary to be able to illuminate a causal mechanism. Table 2 shows the mean itinerary fares for the same hub airports in Table 1 and the mean itinerary fare for non-hub airports over time. Fares are devalued by the US Consumer Price Index into real 1997 values. The graph indicates that within the sample period, fares have decreased by about 30 dollars just between 1997 and the first two quarters of 2001. The trend over time lends support to the hypothesis that decreasing search costs allow passengers to find the lowest fare products that best suit them, as internet usage increased greatly during the sample period.

Table 2: Mean Fares at Hub Airports and Non-Hub Airport between 1997 and 2003

Further information can be collected from Table 3, indicated below, which highlights the changes over time regarding internet usage in the metropolitan statistical areas to which hub airports belong.

Table 3: Internet Usage Rates in Hub Airport MSAs



Table 3 depicts the mean household internet adoption rate among all regions as well as the minimum and maximum for each year. The standard deviation for each year is consistently between 4% and 5%, suggesting that even within regions in one year there is significant cross-sectional variation between MSAs that have hub airports. The data additionally indicates that there is variation over time within unique regions. In 1997 the mean internet adoption rate was 15.46% in all hub airport MSAs, subsequently increasing to a mean of 47.41% by 2003. The sample period was a time when most MSAs were quickly gaining access to the internet, and even though the survey was not collected between 2004 and 2010, there is reason to suspect that there was diminishing growth in internet usage rates after 2003. It is interesting to note that even by 2011 the mean internet rate had barely surpassed 60%.

# Section 5 – Results

The purpose of this section is to detail the results of the model described in Section 4.3, resolve issues of endogeneity as introduced in Section 4.4, and perform a series of robustness checks to check the validity of the findings.

## 5.1 Models 1 and 2

Table 4, on the following page, lists the first two models of this research. The purpose of each model is to compare the introduction of as an omitted variable and record the effect on the coefficient of *AirportShare*. The first model is estimated with OLS without carrier, route, or time fixed effects. The coefficient of *AirportShare* is both positive and significant at the 1% level. It indicates that an increase of one percentage point of *AirportShare* is associated with an increase of the mean itinerary fare of 0.44%, holding all else equal. The results in column two indicate that an increase in the usage of internet in an MSA is associated with a decrease in fare price, but only within certain years. Interestingly, in Model 1 *IntRate* has a positive effect on the average fare, and the effect is significant at the 1% percent level. The dummy variable interactions between *IntRate* and each year indicate that as time progresses, the positive effect of the internet diminishes. How does the variable of interest, *AirportShare,* change? Once controlling for internet usage, the magnitude of the coefficient decreased, but the value of the coefficient on *Rate\*Share* suggests a positive interaction between airport shares and internet usage that is significant at the 1% level. That is, the impact of *AirportShare* now depends on the value of *IntRate*, so that if the rate of internet usage is 30%, a one unit increase in *AirportShare* results in a 0.437% increase in the mean itinerary fare. This is the same increase from column one, so that for values of *IntRate* lower than 30%, the net effect of *AirportShare* is less than the model estimated in column one. The r-squared of Model 1 is relatively low, which should be expected given that fixed effects have not been introduced yet and market characteristics have not been controlled. Model 2 uses the same variables are Model 1 but now includes dummy variables to control for each route-carrier combination and introduces year fixed effects (which are unreported in Table 4.) Column four is specified exactly according to the price equation written in Section 4.3. The output again finds evidence that *AirportShare* increase fare price, although including fixed effects only limits the



increase to 0.39% for every percentage point increase in AirportShare, holding all else equal. A one standard increase of *AirportShare* away from its mean results in an increase of mean itinerary fare of 4.68%. This standard deviation increase translates into a respective fare increase of about $10.06, holding all else equal. Once introducing the measure for internet usage, there is again a decrease in the coefficient of *AirportShare*, and the effect is significant at the 1% significance level. The coefficient of *Rate\*Share* indicates a positive interaction between internet usage and airport shares. The coefficient of *IntRate* in Model 2 is positive and insignificant at the any meaningful level, and the subsequent interactions with year dummy variables imply that internet usage in 2003 could decrease prices compared to 1997. On a final note, a noticeable difference between Model 1 and Model 2 is the vast increase in the R-squared. The coefficients estimated by pooled OLS are able to explain about 10% of the variation in the sample, whereas the fixed effects model captures around 34% of the variation.

## 5.2 Model 3

Model 2 is an obvious improvement from Model 1, as the fixed effects alone explain about an additional 24% of the total variation in *LnFare.* However, there are still significant concerns regarding its validity. As discussed earlier, the variables *Rate\*Share* and *AirportShare* theoretically suffer from simultaneity bias and correlation with the error term. Model 3 corrects for this bias by estimating the price equation with an instrumental variables fixed effects model. This model will be chosen to assess the hypotheses presented in Section 3. Table 5 (on the following page) shows the results of this model. Some concern was initially expressed about the inclusion of *OriginPop* in the model, especially because the endogenous regressors are already predicted by variation of endpoint population data, but the results of an F test showed that *OriginPop* should not be excluded from the model. Column 1 shows the regression results before the addition of *IntRate.* The Kleibergen-Paap rk LM statistic is above 10,000 and the associated p-value for this test is zero, indicating that the set of instruments as outlined in Section 4.4 are relevant to the endogenous variables. Additionally, the Kleibergen-Paap rk Wald F statistic is 576 which is not only above the standard rule of thumb value of 10, but also well above the computed Stock-Yogo critical values, indicating the instruments perform well and do not lead to bias. The Hansen-J statistic rejects the null hypothesis at a 1% significance level, indicating that the excluded instruments are valid and exogenous, and an LM test of redundancy for each instrument shows



that they each improve the asymptotic efficiency of the IV estimation. Lastly, an endogeneity test rejects the null hypothesis at a 1% significance level, meaning *AirportShare,* and *Rate\*Share* can indeed be assumed endogenous. The results of these tests are nearly identical for the results in column six, and the same conclusions can be drawn from each test.

In Model 3, the coefficients of *AirportShare* are positive and significant at the 1% level both before and after the introduction of *IntRate.* Examining column one, we find that a one percent increase in carrier *i’s* share of departing passengers in the market endpoint airports of route *k* increases the average price by 0.669%. In column six, after the introduction of *IntRate,* the coefficient increases to 0.716%. However, the effect of *AirportShare* cannot be disentangled from the effect of increased internet usage. If the value of *IntRate* is 20%, we can expect that a one percentage point increase in *AirportShare* increases the mean itinerary fare by 0.676%. If the value of *IntRate* rises to 45%, then the subsequent fare increase is 0.626% for each one percent increase in *AirportShare.* In simpler terms, a one standard deviation increase in *AirportShare* raises the mean fare by 8.11% if the internet usage rate is 20%, whereas the mean fare is increased by 7.51% if the internet usage rate is 45%. This difference of low and high internet results in respective fare increases (evaluated at the mean fare of $215) of $17.44 and $16.15. This is slightly lower than the amount ($18) found by Evans and Kessides (1993), although their data is from 1988 and direct comparison may not be applicable. The decrease in the dominance premium, illustrated by the negative and significant coefficient on *Rate\*Share,* proves Hypothesis 1 true. That is, once controlling for higher rates of internet usage the pricing premium of airport dominance decreases. Furthermore, the positive and significant coefficient on *AirportShare* proves Hypothesis 2 true, so that each additional percentage increase in the share of passengers at both endpoints of a route will increase a carrier’s average fare.

The results in column two show that a firm true or false cannot be awarded to Hypothesis 3. The model suggests that low values of internet usage in the years 1997-2000 increase the mean fare, and only until the years 2001-2003 does the net effect of *IntRate* become negative. This is because the specified model indicates that the effect of increased internet usage depends on both the value of *AirportShare* and the year of the sample. For example, assuming a constant value of 32% for *AirportShare* (the mean percentage for hub airports), the effect on the mean airfare for a 1% increase in household internet usage in 1997 is a price increase of 0.13% (about 28 cents per percentage point increase in *IntRate*). In 2000 the effect is reduced to an increase in price of 0.004% (less than 1 cent per percentage point increase in *IntRate*), and in 2003 the effect is further reduced to a price decrease of 0.15% (about 32 cents per percentage point increase in *IntRate*).

It should be noted once again that the variables *LoadFactor, RouteShare,* and *LCCroutes* are not instrumented and remain endogenous. We therefore refrain from interpretation of these variables.

## 5.3 Extensions

To check the validity of the findings in the previous section, alternate specifications have been considered and the results are presented below. Referenced tables are available in Appendix 2.

### 5.3.1 Extension 1

As previously mentioned, the coefficients of the endogenous control variables should not be interpreted because they are biased and inconsistent. Covariance between *AirportShare* and *Rate\*Share* with these control variables means that the inconsistency (but not the bias) is also smeared across the coefficients of the variables of interest. Of additional concern is the relatively strong correlation (0.75) between *AirportShare* and *RouteShare*. This could potentially impede clear interpretation of the results, as a one unit increase in one variable cannot be assumed unrelated to the other. That is, we cannot say for certain that a one percentage point increase in *AirportShare* results in a increase in the mean fare ceteris paribus. Table 6 estimates the equation used in Model 3 but instruments for every control variable and drops the variable *RouteShare.* The coefficients in Model 4 are therefore consistent, but the exclusion of *RouteShare* introduces omitted variable bias. Because we know that it is positively correlated with *AirportShare* and *LnFare,* the coefficients on *AirportShare* and *Rate\*Share* will have an upwards bias, suggesting an upper boundary for the true effect. The results in Table 6 indicate that the magnitude of *AirportShare* is slightly larger with an estimated coefficient of 0.0088 (compared to 0.0072 in Model 3). The effect of *Rate\*Share* has doubled in magnitude and the effect of increased internet usage still only decreases the mean fare between the years 2001-2003.

### 5.3.2 Extension 2

Table 7 re-estimates the IV fixed effects model as used in Model 3 on another sample that includes all economy class fares from all airlines in only non-hub airports. The change in *AirportShare* for the new sample is rather striking. In hub airports, an additional percentage point increase of theshare of departing passengers in the market endpoint airports increases the average fare by 0.72%, whereas in non-hub airports it only increases mean fares by about 0.24%. Additionally, the effect of increased internet usage is significant and has a negative effect on mean fares in all years of the sample, not just between 2001-2003. This is most likely due to the fact that, as Orlov (2011) finds, the effect of the internet is strongest in competitive routes. More of the routes out of non-hub airports may be competitive and therefore the regression is estimating a stronger effect for the internet.

### 5.3.3 Extension 3

This thesis has examined the effect of airport dominance and the internet on the average route fare, but perhaps the story changes for fares at the 20th and 80th percentile fares along a route. This is common practice in the literature, as it is conceivable (based on what is known from airline revenue management practices) that a larger share of the 80th percentile fares belongs to business class passengers while the opposite is true for 20th percentile fares. As mentioned in Section 2.4, it is fairly certain that leisure passengers benefit from increased price transparency, but it is probable that business passengers also benefit from increased internet usage as well. This is due to the fact that, as may be more likely for small and medium sized firms, travel policies restrict the purchasing behavior of employees. The results in Table 8 respectively show the results of the IV fixed effects model estimated on 20th and 80th percentile route fares. Interestingly, the effect of *IntRate* and *Rate\*Share* are insignificant for fares at the 20th percentile, whereas the coefficients for both variables are significant at the 80th percentile fare. This suggests that in hub airports, passengers who purchase cheaper fares do not benefit from internet purchases. It may be the case that the dominant airline only offers a certain number of low-price products and lower competition makes the search effort irrelevant. Conversely, fares in the 80th percentile in hub airports are negatively influenced by internet usage between the years 2001-2003. Additionally, the magnitude of the coefficient of the interaction term between *AirportShare* and *IntRate* is larger for the 80th percentile fare than the mean fare.

# Section 6 – Concluding Remarks and Discussion

The purpose of any good conclusion is to synthesize and reflect on observations, yielding new insights; not to simply summarize. Was there any purpose in conducting this research and did it generate any practical value or contribution? When beginning this thesis, awareness in the possibility of a declining hub premium directed us to consider the factors involved that allow airlines to maintain market power, and to examine the ways in which these things have changed. Past research has focused mainly on airport facilities or frequent flyer programs and changes in these elements which affect hub premiums. Yet these are only half of the story. Airline access to and control over computer reservation systems remained just as strong during the late 90’s and early 2000’s, which is the sample period most researchers were using when studying the hub premium. Borenstein (1989) even wrote that computer systems and potential bias from them and travel agents could lead to suboptimal outcomes for passengers. With ample scientific and anecdotal evidence that the internet was leading to lower fares for passengers, it was only natural for us to extend this research to consider if the internet was a moderating effect on hub premiums. It is conceivable that passengers who book their tickets on the internet have reduced search costs and increased price transparency, and therefore are able to more easily choose lower fares or compare prices between airlines. Brunger (2010) shows that even when identical fares are available, passengers who book on the internet purchase fares which are more than 4% lower than passengers who book with travel agents.

The results of this thesis contrast slightly with the results of Orlov (2011). Rather than a homogenous effect of the internet across all years, we find that the effect also depends on each year of the sample. Extension 2 finds that the internet has an increasingly negative (and consistently negative) effect on the mean fare for non-hub airports, but the main results of this thesis shown in Table 5 suggest that the direct effect of the internet in reducing mean fares is only significant and negative for the years 2001-2003 and that increased internet usage in earlier years unexpectedly results in higher mean fares. The higher fares in earlier years may result from a lack of airline effort to publish fares online, lower consumer trust or familiarity with online price comparison tools, or perhaps because in earlier years the internet could have been seen as a household luxury. If wealthier people in an MSA acquired the internet first, they might be less motivated to compare airfares, or instead used the internet to choose among products within an airline. This last point relates to the paper by Granados, Gupta, and Kauffman (2012), who find that product discrimination leads to more price inelastic passengers.

The purpose of this thesis was to illustrate how increased usage of the internet could moderate the effect of airport dominance in hub airports. Without controlling for any effect of the internet (as seen in column one of Table 5) we find that each additional percentage point increase in the share of departing passengers increases the mean fare by 0.669%. Once controlling for the internet, we find that lower levels of internet usage are associated with a higher dominance premium and that as the internet usage rate increases, the dominance premium in hubs decreases. Specifically, internet usage rates below 24% result in a dominance premium higher than 0.669%. When the internet usage rate is 47%, the mean usage rate in hub airport MSAs for the year 2003, the dominance premium is decreased to 0.622%. In other words, assuming a one standard deviation increase of the shares of passengers at both route endpoints and a fare of $215, a hubbing airline has a decreased dominance premium of about $1.50 with higher internet usage compared to low internet usage. While technically a statistically significant decrease, the effect is impractical and unsubstantial. It turns out that the internet is not much of a moderating effect on the dominance premium. Are these results believable? In other words, why would the internet not affect a dominant airline’s premium as much as we previously thought? It is not easy to answer this, as the correlation between internet usage and a carrier’s own share of passengers at an originating airport varies widely. For example, the correlation between *IntRate* and the share of originating passengers for Delta at its Atlanta hub is -0.687, while the correlation between *IntRate* and the share of originating passengers for Delta at its Cincinnati hub is 0.494. It may be the case that hub premiums are the result of a unique constellation of factors for an airport and therefore cannot be generalized over a set of airports. The contribution of this thesis to the literature is demonstrating that the hub premium is most likely a result of competition and access to airport facilities and not computer system and travel agent bias.

A major limitation of this thesis is in the variable *IntRate* itself, as it only proxies for the effect of booking on the internet. This lack of precision may explain why we could not always find a significant influence of the internet in all years. Such detailed consumer data is available only through airlines, who may be reticent in publicizing sensitive information. Future research could consider collaborating with an airline for a case study, but the tradeoff is that what you gain in information you lose in external validity. Furthermore, the measure of airport dominance in the hub airports captures the effect of increased shares for all airlines, but it is conceivable that the premium is only afforded to the hubbing airline and not to the remaining airlines. Future research could isolate the effect with an interaction between airport dominance and each carrier. This thesis does not close the book on research into the hub premium and airline dominance; but if anything, it shows that computer system bias, travel agent bias, and the moderating effects of the internet are not chapters that belong inside.

# Appendix 1: Summary Statistics and Correlation Matrix





# Appendix 2: Miscellaneous Tables







# References

Alamdari, F. (2002). Regional development in airlines and travel agents relationship. *Journal of Air Transport Management, 8*, 339–348.

Berry, S., & Jia, P. (2008). Tracing the Woes: An Empirical Analysis of the Airline Industry. *NBER Working Paper No. 14503*, 1-31.

Borenstein, S. (1989). Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry. *The RAND Journal of Economics, Vol. 20, No. 3*, 344-365.

Borenstein, S. (2005). *U.S. Domestic Airline Pricing, 1995-2004.* Working Paper No. CPC05-48, University of California, Berkeley.

Borenstein, S. (2012). *What Happened to Airline Market Power?* University of California, Berkeley.

Brown, J., & Goolsbee, A. (2002). Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy, 110*(3), 481-507.

Brueckner, J., Dyer, N., & Spiller, P. (1992). Fare Determination in Airline Hub-and-Spoke Networks. *The RAND Journal of Economics, 23*(3), 309-333.

Brueckner, J., Lee, D., & Singer, E. (2013). Airline competition and domestic US airfares: A comprehensive reappraisal. *Economics of Transportation Volume 2 Issue 1*, 1-17.

Brunger, W. (2010, November 6). The impact of the Internet on airline fares: The ‘Internet Price Effect’. *Journal of Revenue and Pricing Management, 9*, 66-93.

Castillo-Manzano, J., & Lopez-Valpuesta, L. (2010). The decline of the traditional travel agent model. *Transportation Research Part E: Logistics and Transportation Review, 46*(5), 639-649.

Ciliberto, F., & Williams, J. (2010). Limited Access to Airport Facilities and Market Power in the Airline Industry. *The Journal of Law & Economics , Vol. 53, No. 3*, 467-495.

Douglas, A., & Lubbe, B. (2008). Violation of the Corporate Travel Policy: An Exploration of Underlying Value-Related Factors. *Journal of Business Ethics, 84*, 97-111.

Escobar-Rodríguez, T., & Carvajal-Trujillo, E. (2014). Online purchasing tickets for low cost carriers: An application of the unified theory of acceptance and use of technology (UTAUT) model. *Tourism Management, 43*, 70-88.

Evans, W., & Kessides, I. (1993). Localized Market Power in the U.S. Airline Industry. *The Review of Economics and Statistics, Vol. 75, No. 1*, 66-75.

Flood, S., King, M., Rodgers, R., Ruggles, S., & Warren, J. Integrated Public Use Microdata Series, Current Population Survey: Version 6.0 [dataset].

Minneapolis, MN: IPUMS, 2018.   
<https://doi.org/10.18128/D030.V6.0>

Granados, N., Gupta, A., & Kauffman, R. (2012). Online and Offline Demand and Price Elasticities: Evidence from the Air Travel Industry. *Information Systems Research, 23*, 164-181.

Huse, C., & Evangelho, F. (2007). Investigating business traveller heterogeneity: Low-costvs full-service airline users? *Transportation Research Part E*, 259-268.

Lane, L., & Verlinda, J. (2004). *The Effect of the Internet on Pricing in the Airline Industry.* University of California, Irvine, Department of Economics.

Lederman, M. (2008). Are Frequent-Flyer Programs a Cause of the “Hub Premium”? *Journal of Economics & Management Strategy, Volume 17, Number 1*, 35-66.

Lee, D., & Luengo-Prado, M. J. (2005). The Impact of Passenger Mix on Reported ‘‘Hub Premiums’’ in the U.S. Airline Industry. *Southern Economic Journal, 72*(2), 372–394.

Mason, K. (2000). The propensity of business travellers to use low cost airlines. *Journal of Transport Geography, 8*, 107-119.

Mason, K. (2001). Marketing low-cost airline services to business travellers. *Journal of Air Transport Management, 7*, 103-109.

Minick, A. (2000). Computer Reservations Systems, Airlines, and the Internet. *Journal of Air Law and Commerce, 65*(4), 891-913.

Orlov, E. (2011). How Does the Internet Influence Price Dispersion? Evidence from the Airline Industry. *The Journal of Industrial Economics, 59*(1), 21-37.

Ravich, T. (2004). Deregulation of the Airline Computer Reservation Systems (CRS) Industry. *Journal of Air Law and Commerce, 69*(2), 387-411.

Sengupta, A., & Wiggins, S. (2014). Airline Pricing, Price Dispersion, and Ticket Characteristics on and off the Internet. *American Economic Journal, 6*(1), 272-307.

Snider, C., & Williams, J. (2015, December). BARRIERS TO ENTRY IN THE AIRLINE INDUSTRY: A MULTIDIMENSIONAL REGRESSION-DISCONTINUITY ANALYSIS OF AIR-21. *The Review of Economics and Statistics, 97*(5), 1002-1022.

Stahl, D. (1989). Oligopolistic Pricing with Sequential Consumer Search. *The American Economic Review, 79*(4), 700-712.

1. It should be worth nothing that the methodology of this thesis will be unable to incorporate any dynamic pricing effects, as the data sources used do not indicate how far in advance of the departure date a ticket was purchased. [↑](#footnote-ref-1)
2. The effect of greater consumer information on price dispersion is not monotonic because at one end of the distribution, all consumers have no information and pay the monopolist’s price. When some information is available, dispersion increases as some consumers find lower costs. Eventually when all consumers are perfectly informed, at some point price dispersion again decreases as consumers purchase a single competitive price. [↑](#footnote-ref-2)
3. https://www.transtats.bts.gov/DatabaseInfo.asp?DB\_ID=125&Link=0 [↑](#footnote-ref-3)
4. https://www.transtats.bts.gov/Fields.asp [↑](#footnote-ref-4)
5. https://www.census.gov/data/tables/2000/dec/phc-t-29.html [↑](#footnote-ref-5)
6. All fare classes of low-cost carriers are kept because the Department of Transportation acknowledges that the variable is not harmonized between airlines and fare classes may be reported differently. For instance, during the sample period none of the Southwest flights are recorded as either Economy unrestricted or restricted. Simply eliminating all business class or first class tickets could present a serious sampling issue. [↑](#footnote-ref-6)
7. See Table 1 on page 24 for the list of hub airports, following from Lee and Luengo-Prado (2005) [↑](#footnote-ref-7)
8. Calculated as : [↑](#footnote-ref-8)
9. It should be noted that this decision carries certain consequences which are addressed in further detail in Section 5.3.1 Extension 1. Because the coefficients of *AirportShare* and *Rate\*Share* are now correlated with endogenous control variables, the estimated coefficients are unbiased but also inconsistent. The sample size is large enough that this should not present a problem. [↑](#footnote-ref-9)