

Internet Penetration and Capacity Utilization in the US Airline Industry*

James D. Dana, Jr.
Northeastern University

Eugene Orlov
Compass Lexecon

June 26, 2008

Abstract

Airline capacity utilization, or load factors, increased dramatically between 1993 and 2007, after staying fairly stable for the first 15 years following deregulation. While improvements in demand forecasting, capacity management, and revenue management are all potential explanations, most of these operational changes were made in the 1980's, significantly before the increase in load factors. We argue that consumers' adoption of the Internet, and their use of the Internet to investigate and purchase airline tickets, can explain recent increases in airlines' load factors. Using metropolitan area measures of Internet penetration, we find strong evidence that differences in the rate of change of Internet penetration explain differences in the rate of change of airline airport-pair load factors. We argue that these increases, and a significant part of the associated \$3 billion reduction in airlines' annual costs, represent a previously unmeasured social welfare benefit of the Internet.

* We would like to thank Shane Greenstein, Bruce Meyer, Kathryn Spier and Scott Stern as well as seminar participants at Northwestern, Stanford, MIT, Harvard, Columbia and the IIOC 2008 for extremely helpful comments. Jim Dana is a professor of economics and strategy at Northeastern University. He is also a visiting scholar at Harvard Business School. He can be reached via e-mail at j.dana@neu.edu or jdana@hbs.edu. Eugene Orlov is an economist at Compass Lexecon and can be reached via e-mail at eorlov@compasslexecon.com

1. Introduction

US airline industry domestic passenger capacity utilization, or load factors, have increased from 62% in 1993 to 80% in 2007 after ranging from 57% to 63% in the years since deregulation. One potential explanation for this dramatic increase is the airlines' use of sophisticated revenue management systems. These systems help airlines to forecast demand, more efficiently utilize their aircraft and personnel resources, and create incentives for consumers to choose alternatives to purchasing seats on flights with scarce capacity, even when that capacity was not expected to be scarce. However, revenue management systems were widely adopted in the 1980's, and, therefore, can not explain dramatic increases in load factors in the late 1990's.

In this paper, we argue that the main force responsible for much of the increase in load factors is the rapid increase in consumer Internet penetration, and the associated increase in the use of the Internet as the primary method for investigating and booking airline reservations in the late 1990's and early 2000's. The Internet has given consumers more information about available products including alternative departure times, alternative carriers, alternative airports, alternative legroom, and alternative in-flight durations (the number of stops), which makes it more likely that consumers will take advantage of incentives to travel on flights with excess capacity and more likely that airlines will find it profitable to offer those incentives. The Internet has also increased price competition among airlines which reduces airlines incentives to hold excess capacity. Consistent with this explanation, we find that changes in US metropolitan area Internet penetration rates explain changes in airlines' airport-pair load factors, dramatically reducing airlines' costs.

Until now, research on the economic impact of the Internet has primarily focused on the impact of lower search costs on the level and dispersion of firms' prices. An obvious implication of lower search costs is increased price competition. While the impact on price levels can be dramatic (see, for example, Brynjolfsson and Smith, 2000), the calculated increases in social welfare associated with price decreases are small. We look at the effect of the Internet on a more direct type of allocative efficiency: the improved utilization of

existing goods and services as measured by airline load factors. We find that the elasticity of capacity utilization with respect to Internet penetration is .102 and that the increase in Internet penetration from 1997 to 2003 resulted in an estimated 7.2% increase in load factors or almost \$3 billion in cost savings each year. We argue that at least half of this savings represents a social welfare gain.

Any attempt to measure the impact of the Internet on capacity utilization must address why capacity isn't being fully utilized in the first place. Most economic models assume either spot market pricing, or forward contracts, and conclude that excess capacity arises only when shadow cost of capacity is zero. However, this does not appear to be the main explanation for excess capacity in the airline industry.

Some models in the economics literature, and most in the operations literature, predict capacity may not be fully utilized by introducing price rigidities. Indeed, casual observation suggests airlines typically do not adjust their prices significantly as a departure time approaches and certainly do not set market clearing prices ex post. Instead they set prices in advance and then use sophisticated software to manage the inventory available at each price. Setting prices in advance when demand is unknown can clearly result in allocative inefficiencies and lead to the underutilization of capacity.

We begin by presenting a simple stochastic peak-load pricing model based on Dana (1999a)¹. In the model, airlines set prices prior to knowing the distribution of demand across flights. As in Dana (1999a) airlines offer multiple prices, which induce some consumers to shift their purchases to the off-peak flight even when the firm cannot anticipate which flight is off peak. We then generalize the model by assuming that some customers are fully informed while others observe only the prices for their preferred departure time. We then show that as the fraction of informed consumers increases, airlines' equilibrium capacity falls and airlines' load factors increase. This holds in both competitive and monopoly markets, but effect is strongest when the market is competitive.

¹ Other papers that examine stochastic peak-load pricing are Carlton (1977) and Brown and Johnson (1969), but these papers consider a social planner who is restricted to uniform prices. Dana (1999a) shows that the competitive equilibrium prices in these models are generally non-uniform.

The model predicts that a decrease in market frictions leads to an increase in load factors, an associated decline in capacity utilization, and an unambiguous increase in social welfare.

To test model predictions, we estimate reduced-form regressions of airline directional airport-pair load factors on metropolitan area Internet penetration. Airline load factors are available from the Bureau of Transportation Statistics, and data on annual Internet penetration for 1997 to 2003 come from the Computer Use and Ownership Supplement to the Consumer Population Survey. Additionally, we use the Department of Transportation's Origin and Destination Survey to derive average fares and construct weights. The next section of the paper discusses the related literature in airline pricing and the economics of the Internet. Section 3 presents the theoretical model. Section 4 describes our data. Section 5 describes the estimation, and Section 6 concludes.

2. The Related Literature

The theoretical literature on capacity decisions and demand uncertainty with price rigidities is extensive. The first set of papers in this category is the stochastic peak-load pricing literature (Brown and Johnson, 1969 and Carlton, 1977). In these models, firms choose capacity and set prices for multiple flights before learning demand. After demand is realized, consumers purchase their preferred product subject to availability. Stochastic peak load pricing predicts that capacity will be underutilization at off-peak times because prices are set before demand is realized.

Note that there is less incentive for consumers to switch from a peak flight to an off-peak flight when firms use uniform prices. By using price dispersion, firms can increase demand-shifting. The earliest paper on price dispersion as a response to demand uncertainty is Prescott (1975) who considered a simple competitive model with a single good. Several papers in the industrial organization literature have built on Prescott's work, including Dana (1998, 1999a, and 1999b), and Deneckere, Marvel, and Peck (1997).² In particular, Dana (1999a) shows that price dispersion increases demand shifting and in so doing increases

² The Prescott model has also been widely applied in monetary economics (see Eden, 1990 and 1994, and Lucas and Woodford, 1993) and labor economics (see Weitzman, 1989).

social welfare by improving the allocation of consumers to available capacity.

Few papers have tried to empirically test the Prescott model. One exception is Escobari and Gan (2007) who directly test the hypothesis that price dispersion is induced by demand uncertainty. They also show that airline price dispersion increases with competition as implied by Dana (1999a and 1999b).

Another exception is Puller, Sengupta, and Wiggins (2007). They have detailed data on airline tickets purchased through a single computer reservation system which allows them to ask what portion of fare differences is associated with restrictions and what portion represents pure dispersion of the type predicted by Dana (1999b). They find modest support for Dana (1999b) and strong support for models based on second-degree price discrimination.

While our paper does not directly test the Prescott model, we test an important implication of the theory. Namely, we test the prediction that capacity utilization increases when consumers are better informed about the available products and their prices.

The empirical literature on the impact of the Internet is extensive. Many papers have compared online markets to traditional markets, and in particular, focused on price levels and price dispersion (see Ellison and Ellison, 2006). Brynjolfsson and Smith (2000) report that compact disk and book prices are 9 to 16% lower in online markets and that price dispersion is slightly smaller. It is not immediately apparent whether price differences reflect differences in costs, or differences in margins, but Brynjolfsson and Smith conclude that the significant sources of heterogeneity, such as brand and reputation, are not diminished by Internet competition. Other papers (for example, Clay, Krishnan, and Wolf, 2001, and Baye, Morgan, and Scholten, 2004) have found less evidence of price declines, but all of these papers find consistent evidence that online price dispersion is quite large, even compared to traditional markets.

A handful of papers have considered the impact of the Internet on prices in the airline industry. Clemons, Hann, and Hitt (2002) and Chen (2002) find that prices available from online travel agents are just as dispersed as those available from traditional

offline travel agents. Using national data on Internet use, Verlinda and Lane (2004) find that increased Internet usage is associated with greater differences between restricted and unrestricted fares. Using a cross section of airline tickets purchased both online and offline, Sengupta and Wiggins (2007) find that tickets sold online have lower average prices and that increases in the share of tickets purchased online implies lower *offline* fares and lower price dispersion. Finally, using metropolitan area Internet access and a differences-in-differences estimation strategy similar to ours, Orlov (2007) examines the impact of Internet access on prices and price dispersion in the airline industry. He finds that increases in Internet access are associated with decreases in airport-pair prices. He also finds that increases in Internet access have led to a decrease in interfirm fare dispersion, but an increase in intrafirm fare dispersion.

Several papers have tried to measure other ways in which the Internet increases consumer surplus. Brynjolfsson, Hu, and Smith (2003) show that the Internet enables consumers to obtain hard-to-find books. Ghose, Telang, and Krishnan (2005) argue that the Internet increases the resale value of new products, and Ghose, Smith, and Telang (2006) show that the Internet facilitates the market for used books. Other papers have emphasized that the Internet reduces consumers' offline transportation costs. For example, Forman, Ghose, Goldfarb (2007) conclude that the Internet reduces consumer travel and transportation costs in the market for books.

Undoubtedly, the Internet has also directly impacted firms' costs. For example, the Internet probably helps firms improve their demand forecasts, reduce their communications costs, and more efficiently monitor their workers and suppliers. However, to our knowledge this is the first paper to show that increasing consumer access to the Internet can lower firms' costs.

Our paper is also related to empirical work on inventory management. Gaur et. al. (2005) finds that inventory turns (the cost of goods sold to inventory ratio) are negatively correlated with margins and capital intensity, and positively correlated with unexpected demand (see also Roumiantsev and Netessine, 2006). Gao and Hitt (2007) consider the

impact of information technology on operation decisions, however their focus is on product variety and not on inventory or capacity utilization. Olivares and Cachon (2007) show that competition increases service levels, and hence inventory ratios, in automobile dealerships. Rajagopalan and Malhotraw (2001) document trends in inventory levels and show that finished goods inventories, materials, and work-in-progress ratios have declined in most manufacturing industries, but they do not find the evidence of greater improvements post-1980 as compared to pre-1980. Finally, in the macroeconomics literature Kahn, McConnell, and Perez-Quiros (2002) use firm level data to test the impact of information technology on the volatility of inventories. They find that information technology has led to a reduction in aggregate output and inflation volatility. However they do not directly address a question of how information technology lowers inventory costs.

3. Theory

In this section we present a generalization of the model of stochastic peak load pricing described in Dana (1999a). Dana (1999a) is an attractive starting point because of its simplicity, and because, like other models of stochastic peak load pricing, it predicts that firms will not always fully utilize their capacity in equilibrium. The model also allows us to easily characterize the role of market power on the way in which airline seats are allocated.

Suppose there are two possible departure times, A and B , and that a finite measure N of consumers have heterogeneous departure time preferences and heterogeneous willingness to pay for their preferred departure time. Suppose consumers' valuations for their preferred departure time, V , are identical, but the disutility from traveling at their least preferred time, w , is distributed with cumulative distribution function $F(w)$ and probability density function $f(w)$ satisfying the monotone hazard rate condition (i.e., $F(w)/f(w)$ is strictly increasing in w). Consumers' departure time preferences and the strength of their preferences, w , are assumed to be independently distributed.

Departure time preferences are correlated across consumers and which of the two departure times, A or B , will be most popular is unknown to the firm. We assume either time is equally likely to be the peak and that the number of consumers who prefer the peak

time is N_1 , which is greater than the number who prefer the off-peak time, that is $N_1 = N - N_2 > N_2$.

The cost of making a single seat available for a single departure is k whether or not the seat is filled and the marginal cost of carrying a single passenger, conditional on having an available seat, is assumed to be 0.

Below we separately consider pricing in competitive and monopoly environments.

3.1 Competitive Pricing

The timing is as follows. First, the firms set their capacity. Second, firms set their prices for their capacity at time A , and their prices for their capacity at time B . Third, the state is realized and consumers learn their departure time preferences and w . Fourth, a fraction α of consumers observe all prices and a fraction $1-\alpha$ of consumers observe only the prices for their preferred departure time. Finally, in random order consumers make their purchase decisions maximizing consumer surplus subject to availability (and assuming they cannot purchase a product they don't observe).³

Following Dana (1999a), in a perfectly competitive market, there are two equilibrium prices, $p_L = k$ and $p_H = 2k$, and the number of seats sold at both departure times at the low price is given by⁴

$$Q_L = N_2 + (N_1 - Q_L)\alpha F(k)$$

and the number of high-priced seats available at both departure times, but which sell only at the more popular departure time, is given by

$$Q_H = (N_1 - Q_L)(1 - \alpha F(k)).$$

It follows that

³ As in Dana (1999a) we use random, or proportional, rationing which seems to be more intuitive for the airline industry application than efficient, or parallel, rationing.

⁴ There are exactly two prices in equilibrium because there are just two demand states in this simple stylized model; one can think of the market as supplying two products, seats that are used when demand is either low or high and seats that are used only when demand is high (see Dana, 1999a for a proof that there are only two equilibrium prices).

$$Q_L = N_2 + (N_1 - N_2) \frac{\alpha F(k)}{1 + \alpha F(k)}$$

and

$$Q_H = (N_1 - N_2) \frac{1 - \alpha F(k)}{1 + \alpha F(k)}.$$

Total capacity is

$$Q_H + Q_L = N_2 + (N_1 - N_2) \frac{1}{1 + \alpha F(k)} \quad (1)$$

and the capacity utilization rate (or load factor) is

$$LF = \frac{Q_H + 2Q_L}{2Q_H + 2Q_L} = \frac{N_1 + N_2}{2N_1 \frac{1}{1 + \alpha F(k)} + 2N_2 \frac{\alpha F(k)}{1 + \alpha F(k)}}. \quad (2)$$

We can write the expression for load factor as

$$LF = \frac{1}{1 + \gamma \left(\frac{1}{2} - A \right)}, \quad (3)$$

where

$$\gamma = \frac{N_1 - N_2}{N_1 + N_2}$$

is a measure of the volatility of demand for each departure time, and

$$A = \frac{\alpha F(k)}{1 + \alpha F(k)}$$

is a measure of the propensity of consumers to switch products that depends on Internet penetration, α , costs, k , and the distribution of waiting costs, w . Equations (2) and (3) imply the following:

Proposition 1. *In a competitive market, the equilibrium load factor is decreasing in the level of market frictions, i.e., increasing in α , and the equilibrium capacity is increasing in the level of market frictions, i.e., decreasing in α .*

Notice that social welfare increases as α increases. Some consumers are clearly better off because an increase in α makes them aware of additional products and hence increases their choice set. Also, since in equilibrium more consumers choose to purchase off peak, more of the firm's capacity is filled whether or not the flight is peak, so in equilibrium more low priced seats are available, which implies that when consumers make their purchase decisions more consumers have the option to purchase low priced seats. So an increase in α makes consumers better off whether or not they are one of the consumers who acquires more product information. Since consumers are strictly better off, and since firms continue to earn zero profits, social welfare increases.

Also, notice that an increase in α decreases airlines' capacity and, as a result, airlines' costs. In a competitive equilibrium, airlines still earn zero profits, so these cost savings are passed on to consumers. As α increases, the proportion of consumers who pay $2k$ falls and the proportion who pay k increases.

Corollary. *A lower bound on the social welfare gains from an increase in α is one half of the cost savings associated with the decrease in equilibrium capacity.*

Proof. Increasing α increases the number of consumers who switch from their preferred flight to an alternate flight. Social welfare increases, because for every additional consumer who switches, costs fall by $2k$. The switchers save k themselves, because they pay k instead of $2k$. While these consumers also bear a cost, because they switch voluntarily, it follows that $w < k$. Also for every consumer who switches, one consumer who does not switch pays a lower price, k , instead of $2k$. So under random rationing, welfare increases by $2k - E[w|w < k] > k$. The welfare increase (per switcher) is strictly greater than one half the cost savings (per switcher). □

3.2 Monopoly Pricing

Now consider the monopolist's pricing problem. Following Dana (1999a), the monopolist offers at most two prices, p_H and p_L . Clearly $p_H = V$, so the monopolist's problem is to choose p_L , or equivalently the discount $d = V - p_L$, to maximize its profits where

$$Q_L(d) = N_2 + (N_1 - N_2) \frac{\alpha F(d)}{1 + \alpha F(d)}$$

and

$$Q_H(d) = (N_1 - N_2) \frac{1 - \alpha F(d)}{1 + \alpha F(d)}.$$

As in (2), we can write the monopolist's load factor as

$$LF = \frac{1}{1 + \gamma \left(\frac{1}{2} - A_m \right)},$$

where

$$\gamma = \frac{N_1 - N_2}{N_1 + N_2}$$

and

$$A_m = \frac{\alpha F(d)}{1 + \alpha F(d)},$$

however, unlike (2), A_m now depends on d , the monopolist's pricing decision.

The monopolist maximizes

$$\max_d 2Q_L(d)(V - d - k) + Q_H(d)(V - 2k).$$

The first-order condition is

$$-2 \left(N_2 + (N_1 - N_2) \frac{\alpha F(d)}{1 + \alpha F(d)} \right) + 2 \frac{\alpha f(d)(N_1 - N_2)}{(1 + \alpha F(d))^2} (k - d) = 0$$

or

$$- \left(N_2 + (N_1 - N_2) \frac{F(d)}{1 + \alpha F(d)} \right) + (N_1 - N_2) \frac{f(d)}{(1 + \alpha F(d))^2} (k - d) = 0. \quad (4)$$

When $d = k$ the left-hand side of the first-order condition is negative, so clearly $d < k$ and $A_m < A$. That is, at the monopoly prices, fewer customers shift their purchases from the peak to the off-peak flight than in a competitive market. This implies

Proposition 2. *All else equal, load factors are lower in a monopoly market than in a competitive market.*

However, just as in the case of competitive markets, the monopolist's load factor rises and capacity falls as α rises. Holding d fixed, it is clear from the definitions of Q_L and Q_H that this is true, and (4) implies that $\frac{dd}{d\alpha} > 0$ so increasing α induces even more switching. Thus, we have:

Proposition 3. *In a monopoly market, the equilibrium load factor is decreasing in the level of market frictions, i.e., increasing in α , and the equilibrium capacity is increasing in the level of market frictions, i.e., decreasing in α .*

3.3 Discussion

The competitive model depicts load factor as depending on two characteristics of a route: the variation in demand, and the propensity of consumers to switch from the peak to the off-peak departure. While we describe the propensity of consumers to switch as a single characteristic, it is a function of the information available to consumers, α , the strength of their preferences (the distribution of w), and the costs, k . Load factors are higher when the cost of an empty seat is higher (longer flights), consumers waiting costs are lower (more leisure travelers), and consumers have more information.

Variation in demand is the most important driver of load factors. We cannot observe demand uncertainty directly, but obvious source of variation in aggregate demand uncertainty is size of the market (i.e., the number of passengers). In a simple environment in which the aggregate demand distribution is the sum of independent binomial decisions, it is

well known that the aggregate demand will be approximately normally distributed with a mean proportional to the number of consumers and a standard deviation proportional to the square root of the number of consumers. So, as the number of consumers grows, the average load factor for an airline that sets its capacity equal to a fixed multiple of mean demand is clearly increasing. Even when it chooses its capacity optimally, the ratio of expected sales to capacity will grow with the size of the market.

An important empirical question is whether the appropriate measure of size is a market measure or a firm measure. If products are perfect substitutes (as in the model above) market size might be more appropriate measure. However, when airlines' products are differentiated, firm size may be a more appropriate control.

A comparison of the monopoly model and the competitive model suggests that market structure is also an important determinant of load factor. In addition, when firms have market power, the equilibrium prices and load factors will also depend on characteristics of consumer demand.

Finally, our model is quite stylized and does not capture important sources of variation in the airline industry that affect equilibrium capacity utilization. For example, the hub and spoke system is likely to increase load factors. By increasing density on its spokes, airlines are able to increase frequency and take advantage of size to reduce the demand uncertainty. Other complex network scheduling decisions will also impact an airline's capacity utilization. For example, an airline may schedule one of its larger planes to fly late in the evening (typically off-peak), so that it is available at its hub in the morning (typically peak). These network scheduling problems are even more complex than it first appears because of legal and union constraints on flight crews' daily flying hours.

With these caveats in mind we now bring the model predictions to the data.

4. Data

We use three different data sets. First, we use the T100 (Form 41) database from the Bureau of Transportation Statistics. This dataset reports the monthly capacity and

passenger traffic by airline, by directional airport-pair *segment*, and by aircraft type, for all the domestic passenger flights in the US. A directional airport-pair flight is a single take-off and landing by a single airplane traveling from one airport to another. The airport-pair-airline unit sales in the T100 database are used to calculate each airlines' market share in each directional airport-pair segment. The capacity and passenger data are also used to calculate the average load factor for each airline in each airport-pair segment. We aggregate the data to the quarterly level.

Second, we use the Computer Use and Ownership Supplement to the Consumer Population Survey (CPS) to measure Internet penetration for every major metropolitan statistical area (MSA). The survey asks about Internet access at home, school, and business. For each metropolitan area we compute the fraction of respondents answering yes to any of these Internet access questions using sample weights provided by the CPS. The data are available for the years 1997, 1998, 2000, 2001, and 2003, and we interpolate the penetration for years 1999 and 2002. Table 1 provides descriptive statistics for this variable.

Finally, we use the Origin and Destination Survey (DB1B) database. This is a 10% sample of all passenger domestic itineraries purchased in each quarter for each year in our sample (1997 to 2003) and includes the airline, the quarter in which the ticket was used, the itinerary origin, destination and stop-over airports, the fare, and the number of passengers paying the fare. The DB1B database includes two entries for each roundtrip ticket and just one entry for each one-way ticket. That is, a market is defined by the passenger's origin and destination. Importantly, the database identifies which entries are the outbound and return portions of round-trip tickets, so the database also allows us to identify the ticket origin, that is, where passengers start their travel when they fly round-trip.⁵

For simplicity we restrict the DB1B database to itineraries with at most one stop on each directional market. We also dropped itineraries where one of the carriers on any segment was unknown, itineraries with "top-coded" fares, and itineraries with fares below \$25 in 2000 dollars. We also dropped very short trips, with travel distance less than 50 miles.

We use the DB1B dataset for several purposes. First, we calculate the average fare for each directional airline-airport-pair segment. To do this, we divide the fare paid for each itinerary among the itinerary's segments in proportion to the distance flown, and then we average fares across passengers who flew that airline-airport-pair segment. Note that because the fare is allocated by distance flown, this is an imperfect measure of the actual incremental cost to consumers of flying on the segment.

Second, we use the DB1B dataset to match Internet penetration to individual passengers flying on each airline-airport-pair segment. Clearly, passengers on a particular flight do not necessarily purchase their tickets in the city that is the flight's point of origin. Most notably, many passengers are returning home on the return portion of a round-trip ticket, so passengers are just as likely to have purchased the ticket in the city that is the flight's destination. Still other passengers will be flying on connecting flights from an origination airport that was different than the airport where the airplane originated and/or to a final destination that is different than the airport where the plane lands. The distinction is important, because our hypothesis is that the level of Internet access where passengers book their tickets, not where the plane originates, is what affects how much information passengers have.

For this reason, we use the DB1B database to identify the true point of origin for each passenger on each airline-airport-pair segment, i.e., the airport at which their itinerary originated. Then using the Internet penetration at each airport's MSA, we obtain a measure of Internet penetration for each passenger. Finally, averaging this measure across passengers (or equivalently, averaging across points of origin using the number of passengers from each point of origin as weights) gives us a measure of Internet penetration which is customized to each airline-airport-pair segment.

For example, consider airline X's flight from airport A to airport B. Assume that 40 percent of the passengers on the flight are flying round trip from A to B, so they are on the outbound leg of their journey; another 35 percent of the passengers on this flight are flying

⁵ Note that before 1999, Southwest Airlines reported all of its roundtrip ticket sales as two one-way tickets, so

round trip from B to A, so they are on the return leg of their journey; another 15 percent of the passengers on this flight are flying round trip from airport F to airport B (with a stop each way in airport A), so this is the second segment in the outbound portion of their journey; finally, assume that the remaining 10 percent are flying round trip from airport A to airport G (with a stop each way in airport B), so they are on the first segment of the outbound portion of their journey. Then the weighted average Internet penetration for passengers on this particular segment is equal to $0.5 I_A + 0.35 I_B + 0.15 I_F$, where I_X is Internet penetration at airport X's MSA.

To control for demographic and economic environment at passengers' MSAs that may be spuriously correlated with Internet penetration, we follow a similar procedure to construct weighted average measures of employment, population and income per capita. These variables at the MSA level are available from the Bureau of Economic Analysis.

After matching these datasets, we further limit our sample to traffic on the 20 largest airlines and between the 75 largest airports in the US. These 20 airlines are listed in Table 2. We exclude Southwest Airlines because of the reporting issues noted above (see footnote 5). We also removed the 3rd and 4th quarters of 2001 from our sample because of the terrorist attacks on 9/11/2001 which severely disrupted service and air travel in that quarter. This leaves us with 87,192 quarterly observations. Table 3 lists descriptive statistics for each of the variables we use in our analysis.

5. Estimation

To test the theory empirically, we estimate a linearized version of equation (4). That is, we estimate a reduced form regression of capacity utilization on metropolitan area Internet penetration and other metropolitan area and quarter-airline-segment characteristics. However, we have limited data on the volatility of demand, costs, and consumer waiting costs, so in addition we control for these omitted exogenous variables with directional airline-segment and airline-quarter fixed effects. The airline-segment fixed effects control for non time-

one cannot identify the ticket origin for Southwest passengers in the DB1B market database.

varying airline, segment, metropolitan area, and airport characteristics, as well as airline-airport characteristics such as the presence of a hub or local brand loyalty. The airline-quarter fixed effects control for time-varying airline-specific characteristics, such as brand loyalty.

Table 4 contains our first set of regressions. We regress the log of the quarterly, airline, directional, airport-pair (segment) load factor on the log of Internet access and our control demographic and economic variables. As discussed above, the measure of Internet access is specific to each directional segment and each quarter (the weights change quarterly, although Internet changes annually). We use a log-log specification in part because we believe that the impact of an increase in Internet penetration is greatest when the level of Internet penetration is small. That is, the early adopters of the Internet are more likely to be air travelers than the late adopters.

Also, while our unit of observation is an airline, directional, airport-pair quarter, the economic unit of observation that is of interest is capacity. Particularly for making welfare calculations, we would like to put more weight on airport-pairs with more flights and more available seats. For this reason, our regressions are weighted by the number of available seats.

Before interpreting our results, it is useful to begin by asking what are the unobserved sources of variation that we may not be controlling for with our fixed effects? The most obvious are segment specific variation in cost and variation in demand. The latter includes variation in the size of market demand as well as variation in the elasticity of demand and extent of demand uncertainty. Other potential sources of variation are changes over time in market structure, as well as changes in the degree of firm rivalry and the threat of entry.

The results in Column 1 indicate a large and significant effect of the Internet on load factors. In Column 2 we introduce average fare on airline-airport-pair segment as an additional control. We find that higher fares lead to lower load factors, which is consistent with the intuition that holding costs fixed, airlines with higher fares are more willing to hold speculative capacity. While it is clear that the biggest source of variation in fare levels is

likely to be costs, these results are not surprising since the variation in fares over time is less likely to be due to route-specific shifts in costs, and is more likely to be because of segment-specific changes in demand, rival behavior, or the threat of entry.

Adding the fare as a control variable reduces the coefficient on Internet penetration from .102 to .072, and the coefficient is statistically significant. There are two reasons to think that Internet penetration should be correlated with fares and that controlling for fare would reduce the impact due to increased price competition. First, if the only impact of the Internet is to increase price competition and if the lower prices cause airlines to reduce their excess capacity (as would be predicted in a simple inventory model), then controlling for fare should eliminate the impact of the Internet. Second, if the Internet directly increases capacity utilization as suggested by the theory above, then airlines' costs will fall, which will cause fares to fall. The fact that controlling for fares, we still find large positive and significant effect of the Internet on capacity utilization, suggests that in addition to indirect effect on load factors, via increased price competition, Internet penetration also has directly impacted load factors.

We are also interested in whether the impact on the Internet on load factors is a consequence of increased competition or better allocation at existing prices. Note, that market structure defined on a segment may not be fully capturing market power because airport-pairs that are served by a single non-stop carrier will face significant competition from airlines offering connecting service. For this reason, instead of defining market structure variables that reflect market conditions on each particular segment, we choose to construct, for each airline airport-pair segment, the fraction of passengers that travel on a monopoly route, the fraction of passengers that travel on a duopoly route, and the fraction of passengers that travel on a competitive route. We use the DB1B dataset to calculate these variables. A monopoly route is defined as a route on which the largest firm's market share (share of passengers) exceeds 90%. A duopoly route is defined as a non-monopoly route on which the largest two firms' combined share exceeds 90% or the two largest firms' combined market share exceeds 80% and the third largest firm's share is less than 10%. All

other routes are considered to be competitive routes.

While our measures of market competition are limited and endogenous, in Column 3 of Table 4 we estimate a reduced form regression including market structure and market share. The results indicate that load factors are lower in more concentrated markets. Firms with market power are likely to have higher margins, which increases the incentive to hold speculative capacity. While market structure is endogenous, it is likely that much of the variation in market structure over time will be driven by financial conditions of airlines and demand changes in other markets. However, clearly these coefficients must be interpreted cautiously.

We find that load factors are higher for firms with a larger market share within each segment. This is consistent with the intuition that the variance of demand uncertainty falls relative to the mean as the market size grows, so the incentive to hold speculative capacity falls with size. Or more simply put, it is easier to match the number of planes to the size of the market when the market is larger.

In Column 4, we introduce both the average fare on a segment and market structure variables. The results do not change substantially from previous specifications. Finally, Column 5 presents estimates of a regression that includes average fare, market structure, market share, and available seats (capacity) as control variables. While these variables are all endogenous, they are likely to be correlated with other unobserved exogenous variables. Seeing that our results are robust to the inclusion of these endogenous variables makes us more confident that our results are not a consequence of correlation between Internet use and other time-varying market characteristics.

However, we find that load factors are lower for firms with more available seats. We expected the opposite result. Variation in demand or costs (that we don't observe but firms do observe) should move capacity and load factors in the same direction. Instead, we are seeing the effect we would expect to see if seats were exogenous: exogenous increases in capacity reduce load factors. The simplest interpretation of these results is that airlines are slow to adjust capacity in response to shifts in demand. However, while the number of seats is

clearly a measure of market size, it is an airline specific measure. It is likely that these shifts in demand are due to capacity adjustments by rivals.

In the last two columns of Table 4 we interact Internet penetration variables with the percentage of passengers on a segment who travel on monopoly, duopoly and competitive routes. The results show that increases in Internet penetration have a positive effect on load factors on all segments, irrespective of market structure on routes where passengers travel. More importantly, the order of the effects is consistent with the theory predictions: the effect is the smallest on segments on which higher fraction of passengers travel on monopoly routes, followed by segments on which higher fraction of passengers travel on duopoly routes, and then followed by segments on which higher fraction of passengers travel on competitive routes.

In Table 5, we report our second set of regressions. If the markets demand is the sum of identically and independently distributed individual consumer demands, then it will be normally distributed with a variance proportional to the square root of the market size. This implies that the ratio of the variance to the expected demand is also decreasing in market size and in equilibrium the ratio of capacity to expected demand will be decreasing in market size. But if capacity utilization is higher on larger routes, it follows that increases in Internet penetration should have less impact on load factors. We test this by interacting Internet penetration with available seats. We find, as expected, that Internet penetration has a larger effect on smaller markets.

6. Social Welfare

We find that Internet penetration has a positive and statistically significant effect on load factors. Using the results in Column 5 of Table 4, the elasticity of Internet penetration on load factor is 0.102. That is, each percentage point increase in Internet penetration increases load factors by .102%, and a doubling in Internet penetration increases load factors by 7.2% (i.e., $2 \cdot 0.102 - 1$). From a starting point of 69%, this implies load factors would increase to 74.3%. In our sample period, Internet access more than doubled in many cities while load

factors have increased from about 69% to 73%. So the increase in Internet penetration appears to explain all of the increase in airline load factors during our sample period.

US airline industry passenger flying operations and maintenance costs were \$40 billion in 2000, so an increase of 7.2% in load factor represents approximately a 7.2% decrease in these costs or almost \$3 billion in cost savings every year.

Because our Internet variable is a noisy measure of how many consumers are using the Internet to choose flights and/or purchase their tickets, our estimates are likely to understate the total cost savings associated with the Internet. In particular, since we cannot identify perfectly where consumers purchase their tickets, we aren't measuring the impact of the Internet on those passengers who purchase their tickets elsewhere. So it is possible that the cost savings could be even larger. On the other hand, it remains possible that the results in Table 4 are the result of an omitted variable bias. Also, since we are attributing almost all of the growth in load factors to the Internet, it is unlikely the Internet had much more impact.

Another serious concern is that this cost savings need not all represent a welfare gain. Some of these cost savings may have been offset by decreases in consumer surplus as consumers elect to travel at less convenient times. Without estimates of the demand function, we cannot measure the lost in consumer surplus associated with consumers switching departure times. However, we argued in our model (see Section 3 above) that the social welfare gains are at least one half of the cost savings. And, if the inconvenience of flying off-peak, w , is small, the gains could be significantly higher.

However, the model may overstate the welfare gains. First, in a competitive model with no aggregate uncertainty or with market clearing spot prices, the impact of a reduction in market frictions would also be to shift demand, but the impact on welfare would be smaller. For example, in a competitive peak-load pricing model that exhibits some under-utilization of capacity, the off-peak price will be c and the peak-price will be $2k + c$. So the welfare gain for each switcher is $2k - E[w|w < 2k]$ which is strictly positive, but significantly smaller than $2k - E[w|w < k]$. However, this is not consistent with casual

evidence. Airline fares for ex post off-peak flights do not generally equal marginal cost but instead are significantly higher.

Second, increasing load factors also increases congestion on airplanes, which may reduce the quality of consumers' travel experience. On the other hand, if airlines fly fewer planes airports may be less congested, on-time performance may increase, and the quality of consumers's travel experience may rise. We are unable to measure either of these effects.

Finally, we have considered a simple model in which seats are not rationed. In a more general model in which there was some limit on market prices or in which demand is lumpy, a reduction in market frictions could theoretically lead to an increase in rationing. With rationing, it no longer follows that the disutility of consumers who switch must be bounded by the difference in fares.⁶

With these caveats in mind, our estimates imply an ongoing social welfare gain of at least \$1.5 billion every year is the result of the increase in Internet penetration that occurred during just the 1997 to 2003 period.

7. Conclusion

The Internet has clearly made it easier for consumers to become informed about alternatives to their preferred time of departure, carrier, or destination. A customer buying a ticket on an airline's web site, such as United.com, or on a third party travel services web site, such as Expedia.com, selects their itinerary from a much larger set of options than those that are available to a customer making a reservation on the telephone. The increase in consumers' information has helped airlines to reduce their capacity costs, and airlines appear to be well aware of this. On United Airlines' web site even after choosing their itinerary from the wide selection available, a customer is shown yet another set of lower fare options before making their final purchase decision. No doubt United is able to capture some of the surplus created when it induces consumers to switch flights, so it is interesting to note

⁶ By rationing, we mean that no seat is available at any fare and in any face class. So while coach seats are sometimes rationed, business and first class seats are almost always available because airlines typically can use unsold seats in these classes as reward or upgrades at the last minute for their frequent fliers.

that it is United, not Expedia, which offers this feature.

We used differences in Internet penetration across time and metropolitan areas to identify the impact of reductions in market frictions on differences in airlines' capacity utilization rates, or load factors, across time and airport-pair segments. We found that an increase in Internet access in metropolitan areas leads to an increase in load factors on flights flown by passengers whose travel begins in that metropolitan area. That is, a flight's load factor increase faster when passengers traveling on the flight come from cities in which Internet use is increasing faster.

While increases in Internet access have led to increases in airlines' load factors and a decrease in airlines' costs of almost \$1 billion each year, we believe that much of this cost savings has been passed on to consumers through lower prices. This is consistent with the fact that airlines did not see dramatic increases in profits during this period. It is also consistent with the empirical literature (particularly Orlov, 2007, and Sengupta and Wiggins, 2007) which has found that the Internet has significantly reduced average airline prices. However, whether or not the cost savings is passed on to consumers, we have argued that much of this cost savings represents an increase in social welfare.

References:

- Baye, M.R., Morgan, J. and Scholten, P. (2004), "Price Dispersion in the Small and the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics*, 52(4), pp. 463-496.
- Brown, G., Jr. and Johnson, M.B. (1969), "Public Utility Pricing and Output Under Risk," *American Economic Review*, 59(1), pp. 119–128.
- Brynjolfsson, E. and Smith, M. (2000), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46(4), pp. 563–585.
- Brynjolfsson, E., Hu, Y. and Smith, M. (2003), "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety," *Management Science*, 49(11), pp. 1580–1596.
- Chen, S. (2002), "Differential Pricing on the Web: The Case of Online Air Travel Market", mimeo.
- Clay, K., Krishnan, R. and Wolf, E. (2001), "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics*, 49(4), pp. 521-540.
- Clemons, E.K., Hann, I.-H. and Hitt, L. (2002), "Price Dispersion and Differentiation in Online Travel: An Empirical Investigation", *Management Science*, 48(4), pp. 534-549.
- Carlton, D.W. (1977), "Peak-Load Pricing with Stochastic Demand," *American Economic Review*, 67(5), pp. 1006–1010.
- Dana Jr., J.D. (1998), "Advanced-Purchase Discounts and Price Discrimination in Competitive Markets," *Journal of Political Economy*, 106(2), pp. 395–422.
- Dana Jr., J.D. (1999a), "Using Yield Management to Shift Demand when the Peak Time is Unknown," *RAND Journal of Economics*, 30(3), pp. 456–474.
- Dana Jr., J.D. (1999b), "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure," *RAND Journal of Economics*, 30(4), pp. 632–660.
- Deneckere, R., Marvel, H. and Peck, J. (1997), "Demand Uncertainty and Price Maintenance: Markdown as Destructive Competition," *American Economic Review*, 87(4), pp. 619–641.
- Eden, B. (1990), "Marginal Cost Pricing When Spot Markets Are

- Complete,” *Journal of Political Economy*, 98(6), pp. 1293–1306.
- Eden, B. (1994), “The Adjustment of Prices to Monetary Shocks when Trade is Uncertain and Sequential,” *Journal of Political Economy*, 102(3), pp. 493–509.
- Ellison, G. and Ellison, S.F. (2005), “Lessons about Markets from the Internet,” *Journal of Economic Perspectives*, 19(2), pp. 139–158.
- Escobari, D. and Gan, L. (2007), “Price Dispersion under Costly Capacity and Demand Uncertainty,” NBER Working Paper No. W13075.
- Forman, C., Ghose, A. and Goldfarb, A. (2007), “Geography and Electronic Commerce: Measuring Convenience, Selection, and Price,” Proceedings of the INFORMS Conference on Information Systems and Technology (CIST 2007), Seattle.
- Gao, G, and Hitt, L.M. (2007), “IT and Product Variety: Evidence from Panel Data,” University of Pennsylvania, Wharton School, working paper.
- Gaur, V., Fisher, M.L. and Raman A. (2005), “An Econometric Analysis of Inventory Turnover Performance in Retail Services,” *Management Science*, 51(2), pp. 181–194.
- Ghose, A., Smith, M. and Telang, R. (2006), “Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact,” *Information Systems Research*, 17(1), pp. 3–19.
- Ghose, A., Telang, R. and Krishnan, R. (2005), “Impact of Electronic Secondary Markets on Information Goods Supply Chain,” *Journal of MIS*, 22(2), pp. 91–120.
- Kahn, J.A., McConnell, M.M. and Perez-Quiros, G. (2002), “On the Causes of the Increased Stability of the U.S. Economy,” *Economic Policy Review*, 8(1), pp. 183-202.
- Lucas, R. E., Jr. and Woodford M. (1993), “Real Effects of Monetary Shocks in an Economy with Sequential Purchases,” Working Paper # 4250, NBER.
- Olivares, M. and Cachon, G. (2007), “Competing Retailers and Inventory: An Empirical Investigation of U.S. Automobile Dealerships,” working paper.
- Orlov, E. (2007), “How Does the Internet Influence Price Dispersion? Evidence from the Airline Industry,” working paper.

- Prescott, E.C. (1975), "Efficiency of the Natural Rate," *Journal of Political Economy*, 83(6), pp. 1229–1236.
- Puller, S.L., Sengupta, A. and Wiggins, S.N. (2007), "Testing Theories of Price Dispersion and Scarcity Pricing in the Airline Industry," Texas A&M University working paper.
- Rajagopalan, S. and Malhotra, A. (2001) "Have U.S. Manufacturing Inventories Really Decreased? An Empirical Study," *Manufacturing and Service Operations Management*, 3(1), pp. 14–24.
- Roumiantsev, S. and Netessine S. (2007), "What Can Be Learned From Classical Inventory Models: A Cross-Industry Empirical Investigation," *Manufacturing and Service Operations Management*, 9(4), pp. 409-429.
- Sengupta, A. and Wiggins, S.N. (2007) "Airline Pricing, Price Dispersion and Ticket Characteristics On and Off the Internet," Texas A&M University, working paper.
- Verlinda, J.A. and Lane, L. (2004), "The Effect of the Internet on Pricing in the Airline Industry", mimeo.
- Weitzman, M.L. (1989), "A Theory of Wage Dispersion and Job Market Segmentation." *Quarterly Journal of Economics*, 104(1), pp. 121–37.

Table 1. Internet Penetration across Metropolitan Statistical Areas (N=243)

Year	Mean	Std. Dev.	Min	Max
1997	0.194	0.074	0.043	0.489
1998	0.413	0.114	0.103	0.699
1999	0.482	0.098	0.210	0.764
2000	0.551	0.106	0.218	0.829
2001	0.652	0.103	0.222	0.911
2002	0.672	0.091	0.277	0.892
2003	0.692	0.099	0.332	0.913

Source: Computer Use and Ownership Supplement to the Consumer Population Survey.

Note: Data for 1999 and 2002 are interpolated.

Table 2. Differences Across Airlines.

	Avg. Fare	Avg. Load Factor	Avg. Segment Market Share
Air Wisconsin Airlines	77.20	0.681	0.436
AirTran	89.68	0.674	0.263
Alaska Airlines	124.36	0.677	0.629
America West	107.93	0.689	0.492
American Airlines	180.38	0.692	0.627
American Eagle	92.40	0.644	0.444
ATA Airlines	122.67	0.727	0.752
Atlantic Southeast Airlines	96.34	0.674	0.168
Comair	110.28	0.663	0.509
Continental Airlines	173.45	0.712	0.789
Delta Airlines	142.14	0.702	0.727
Frontier Airlines	131.51	0.613	0.214
Horizon Air	79.90	0.660	0.565
JetBlue	120.76	0.823	0.738
Mesaba Airlines	125.73	0.568	0.309
Northwest	162.75	0.692	0.822
Spirit Airlines	113.28	0.764	0.262
Trans World Airlines	138.25	0.684	0.763
United Airlines	180.23	0.702	0.633
US Airways	145.31	0.668	0.837

Notes: Each cell contains (weighted by number of seats) average values over each airline's routes in the sample over 18 quarters. Fares are in 2000 dollars and are reported as half of the round-trip fare. Airlines' market share is measured in number of passengers.

Table 3: Descriptive Statistics (87,192 observations)

Variable	Mean	Std. Dev.	Min	Max
Load Factor	0.673	0.163	0.003	1.000
Market Share	0.584	0.371	0.000	1.000
Percent of Passengers on				
Monopoly Routes	0.102	0.171	0.000	1.000
Duopoly Routes	0.453	0.324	0.000	1.000
Competitive Routes	0.445	0.319	0.000	1.000
Weighted Internet	0.537	0.170	0.053	0.854
Segment Fare (\$)	151.81	78.21	4.83	1389.24
Seats	42167	42880	30	419644

**Table 4. Regression Results
With Airline-Segment and Airline-Quarter Fixed Effects**

Dependent Variable:	LOG (Load Factor)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
LOG (WEIGHTED INTERNET)	.102*** (.021)	.072*** (.018)	.100*** (.021)	.067*** (.018)	.069*** (.018)		
EMPLOYMENT (%)	.399** (.186)	.457*** (.168)	.408** (.185)	.456*** (.167)	.533*** (.167)	.401** (.185)	.524*** (.167)
LOG (INCOME PER CAPITA)	.566*** (.066)	.677*** (.062)	.547*** (.066)	.665*** (.062)	.722*** (.063)	.571*** (.067)	.735*** (.063)
LOG (POPULATION)	.007 (.014)	.002 (.013)	.009 (.014)	.004 (.013)	.008 (.013)	.007 (.014)	.007 (.013)
LOG (WEIGHTED INTERNET) * % PASS. ON MONOP. ROUTES						.084*** (.023)	.056*** (.020)
* % PASS. ON DUOP. ROUTES						.099*** (.021)	.066*** (.018)
* % PASS. ON COMP. ROUTES						.107*** (.021)	.074*** (.018)
LOG (FARE)		-.267*** (.010)		-.277*** (.011)	-.307*** (.011)		-.308*** (.011)
% PASS. ON MONOP. ROUTES			-.038*** (.010)	-.018* (.009)	-.019** (.009)		
% PASS. ON DUOP. ROUTES			-.015*** (.005)	-.010** (.005)	-.007 (.004)		
MktSHARE			.112*** (.013)	.146*** (.011)	.232*** (.013)		.221*** (.013)
LOG (SEATS)					-.086*** (.005)		-.086*** (.005)
Observations	87192	87192	87192	87192	87192	87192	87192

Notes: Standard errors are in parentheses. Stars denote the significance level of coefficients: *** - 1 percent, ** - 5 percent, * - 10 percent. The sample includes flights on segments between top 75 airports and operated by top 20 airlines. Weighted Internet penetration by quarter, directional segment, carrier, is calculated as a weighted (by the number of passengers) measure of Internet penetration in the originating airport for all passengers on the carrier's flights on a directional segment. FARE is the average fare on a corresponding segment, calculated from the O&D market-level data proportionally to the distance of the segment in total itinerary. Southwest Airlines is excluded because they report round-trip tickets as two one-way tickets, which precludes the calculation of our weighted Internet penetration variable.

**Table 5. Regression Results
With Airline-Segment and Airline-Quarter Fixed Effects**

Dependent Variable	LOG (Load Factor)	
	(1)	(2)
LOG (WEIGHTED INTERNET)		
* I (SEATS IN 1 ST QUARTILE)	.065 ^{***} (.019)	.124 ^{***} (.019)
* I (SEATS IN 2 ND QUARTILE)	.060 ^{**} (.018)	.075 ^{***} (.018)
* I (SEATS IN 3 RD QUARTILE)	.064 ^{***} (.018)	.061 ^{***} (.018)
* I (SEATS IN 4 TH QUARTILE)	.081 ^{***} (.018)	.069 ^{***} (.018)
EMPLOYMENT (%)	.483 ^{***} (.167)	.536 ^{***} (.167)
LOG (INCOME PER CAPITA)	.671 ^{***} (.062)	.722 ^{***} (.063)
LOG (POPULATION)	.002 (.013)	.008 (.013)
LOG (FARE)	-.270 ^{***} (.010)	-.309 ^{***} (.011)
% PASS. ON MONOP. ROUTES		-.020 ^{**} (.009)
% PASS. ON DUOP. ROUTES		-.008 [*] (.004)
MktSHARE		.233 ^{***} (.013)
LOG (SEATS)		-.094 ^{***} (.005)
Observations	87192	87192

Notes: Standard errors are in parentheses. Stars denote the significance level of coefficients: *** - 1 percent, ** - 5 percent, * - 10 percent. The sample includes flights on segments between top 75 airports and operated by top 20 airlines. Weighted Internet penetration by quarter, directional segment, carrier, is calculated as a weighted (by the number of passengers) measure of Internet penetration in the originating airport for all passengers on the carrier's flights on a directional segment. FARE is the average fare on a corresponding segment, calculated from the O&D market-level data proportionally to the distance of the segment in total itinerary. Southwest Airlines is excluded because they report round-trip tickets as two one-way tickets, which precludes the calculation of our weighted Internet penetration variable.