Internet Penetration and Capacity Utilization in the US Airline Industry*

James D. Dana, Jr.
Northeastern University

Eugene Orlov
Compass Lexecon

May 5, 2009

Abstract

Airline capacity utilization, or load factors, increased dramatically between 1993 and 2007, after staying fairly level for the first 15 years following deregulation. We argue that consumers' adoption of the Internet, and their use of the Internet to investigate and purchase airline tickets, explains this increase. We find that differences in the rate of change of metropolitan area Internet penetration explain differences in the rate of change of airline-airport-pair load factors. Consistent with our explanation, we also find that, all else equal, changes in Internet penetration have a bigger impact on load factors on flights in more competitive markets and on flights with fewer total passengers. We argue that a significant part of the associated $3 billion reduction in airlines' annual capacity costs, represents a previously unmeasured social welfare benefit of the Internet.

* We would like to thank seminar participants at the Kellogg School of Business, Stanford University School of Business, MIT-Harvard Industrial Organization workshop, and Columbia Business School, and conference participants at the 2008 IIOC and the 2009 AEA meetings for their helpful comments. We particularly want to thank Glenn Ellison, Diego Escobari, Dan Greenfield, Shane Greenstein, Mike Mazzeo, Bruce Meyer, Kathryn Spier, and Scott Stern. James Dana is a professor of economics and strategy at Northeastern University, and 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 a senior economist at Compass Lexecon and can be reached via e-mail at eorlov@compasslexecon.com.

1. Introduction

US airline industry domestic passenger load factors, or capacity utilization, have increased from 62% in 1993 to 80% in 2007 after ranging from 57% to 63% in the years since deregulation. One potential explanation is the use of sophisticated revenue management systems by airlines. These sophisticated data and capacity management 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 hence cannot be a sole explanation for increased capacity utilization in the late 1990's.

Instead, we argue that the rapid increase in consumer Internet penetration in the late 1990's and early 2000's, and the associated increase in the use of the Internet as the primary method for investigating and booking airline reservations, is responsible for most, if not all, of the increase in airlines' load factors. 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 has made 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. Consistent with this explanation, we find strong statistical evidence that changes in airlines' airport-pair load factors are associated with changes in US metropolitan area Internet penetration rates.

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 increases in social welfare associated with price decreases can be small.[1] We look at the effect of the Internet on a more direct type of allocative efficiency: the increased utilization of existing

[1] Holding market structure and costs fixed, the welfare gain from the lower equilibrium prices is equal to the reduction in the dead weight loss associated with the firms' mark-ups, significantly less than the increase in consumer surplus, while the welfare gain from a reduction in firms' costs holding price fixed is equal to the entire increase in producer surplus.

resources, 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. Many economic models assume either spot market pricing or forward contracts and conclude that excess capacity exists only when shadow cost of capacity is zero. Other models, including the stochastic peak-load pricing research in the economics literature and most of the revenue management research in the operations research literature, predict that capacity is not 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 ex ante before demand is known clearly result in allocative inefficiencies and lead to the underutilization of capacity.

The empirically testable predictions of increases in Internet use are very similar for these two classes of models. Because we think it is somewhat more descriptive of the airline industry, we present a simple stochastic peak-load pricing model based on Dana (1999a).[2] In our model, airlines set prices ex ante before learning the distribution of demand across flights. As in Dana (1999a) airlines offer multiple prices inducing some consumers to shift their purchases from the peak 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. The model predicts that an increase in the amount of price and product information (e.g., because of increased access the Internet) leads to an increase in load factors, an associated decline in capacity, and an unambiguous increase in social welfare.

[2] 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.

This holds in both competitive and monopoly markets, but effect is strongest when the market is competitive.

We test our theoretical predictions by estimating a reduced-form regression of load factors on metropolitan area Internet penetration. We match the metropolitan area Internet penetration to the segment load factors by measuring the fraction of passengers on a given segment whose travel originates in each metropolitan area, since this is the metropolitan are where the tickets were most likely purchased. Also, by using airline-quarter and airline-segment fixed effects we are able to identify the impact of the Internet on load factors controlling for unobserved market, airport, airline, and time-specific characteristics. That is, we test whether differences in the rate of change of Internet penetration in the metropolitan area where an airline's passengers' travel originate can explain differences in the rate of change of airline-airport-pair load factors.

We estimate that the elasticity of load factors with respect to Internet penetration is about 0.107, and hence a 100% increase in Internet penetration implies load factors should increase by over 7%, which is roughly consistent with the observed increase in Internet use and load factors during our sample period. We also find that load factors are higher in more competitive markets and that the impact of Internet penetration is greater on flights in more competitive markets and greater on flights serving lower volume airport-pair segments, which is consistent with our theoretical model and more generally consistent with our theory that the Internet has increased load factors by increasing consumers' price and product information.

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, while Section 4 describes our data. Section 5 describes the estimation and results, and Section 6 concludes.

2. The Related Literature

The traditional peak load pricing literature, see in particular Boiteaux (1980), assumes prices are used ex post, once demand is known, to shift demand from peak to off-peak times. This is desirable either because capacity is set before demand is known, or constrained to be identical at the peak and off-peak times. In the stochastic peak-load pricing literature (Brown and Johnson, 1969, and Carlton, 1977), firms choose capacity and set prices for the different times ex ante, before learning consumer demand. After demand is realized, consumers purchase their preferred product subject to availability. Stochastic peak load pricing predicts that capacity will be underutilized at off-peak times because prices are set before demand is realized. Crew and Kleindorfer (1986) review both literatures.

Note that there is little incentive for consumers to switch from a peak flight to an off-peak flight when firms set uniform ex ante prices. However 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). In particular, Dana (1999a) shows that price dispersion increases demand shifting and in so doing increases 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 are associated with restrictions and what portion represent pure dispersion of the type predicted by Dana (1999b). They find modest support for Dana (1999b) and strong support for models of second-degree price discrimination.

The empirical literature on the impact of the Internet is extensive. Many papers have compared online markets to traditional markets, and in particularly, 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 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 imply 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 fares and 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 and 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 helps firms improve their demand forecasts, reduce their communications costs, and more efficiently monitor their workers and suppliers. However, to our knowledge this paper is the first paper to show that increasing consumer access to the Internet lowers firms' costs.

Our paper is also related to empirical work on inventory management. Gaur, Fisher, and Raman (2005) find that inventory turns (the cost of goods sold to inventory ration) 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. Cachon and Olivares (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 that 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

We consider a model of stochastic peak load pricing, based on Dana (1999a), in which firms set both capacities and prices ex ante, before they or consumers learn consumer demand. The model generates clear, testable predictions for the impact of market power and consumer information on the equilibrium capacity utilization. As we discuss later, these testable predictions are also common to a variety of alternative specifications of the model, including specifications in which capacity is set ex ante, but prices are set ex post after firms learn demand.

Assume two departure times, *A* and *B,* each of which is equally likely to be the peak period, and assume the fraction of consumers who prefer the peak departure time is $\phi_H > \phi_L$ where $\phi_H + \phi_L = 1$. Assume *N* consumers have uniform valuations *V* for their preferred departure time and $V - w$ for the other departure time. Assume the distribution of the disutility from traveling at their least preferred time, *w,* is $F(w)$. We interpret *w* as the cost of waiting for the next flight. Finally, assume that a fraction α of all consumers are informed and observe all of the firms' products and prices and that a fraction $1-\alpha$ of all consumers observe only the prices for products offered at their preferred departure time.[3,4] We interpret α as a measure of the share of consumers with access to the Internet for price and product information.

The cost of capacity is k per seat departure plus c per passenger departure. So a firm with q units of capacity and $q/2$ passengers at each of its two departure times has costs $2kq + cq$.

Firms choose their capacity and the price for each unit of their capacity ex ante before they know which departure time is the peak time. The exact timing is as follows:

1) The firm or firms choose their capacity
2) Firms set their prices for their capacity for each departure time.
3) Consumers learn their departure preferences, w.
4) Informed consumers observe all of the firms' prices and products. Uninformed consumers observe only the prices and products at their preferred departure time.
5) Consumers purchase in random order given availability and the restriction that the uninformed consumers cannot purchase a product that they do not observe.

For example, suppose that firms offer $\phi_H N$ units of capacity at each departure time and sell all of their capacity at a uniform price, V. Then peak period sales are $\phi_H N$, off-peak period sales are $\phi_L N$, and the average capacity utilization rate is $1/(2\phi_H) < 1$. In this example, each consumer flies at his or her preferred time because peak capacity is not rationed and the price of the off-peak departure time is the same as the price of the peak departure time, so there is no benefit from switching.

On the other hand, if firms sell their capacity at multiple prices at each departure time, then once enough consumers have made their purchase decisions, the remaining consumers will face higher prices at the peak departure time then at the off-peak departure

[3] For simplicity, we assume that each consumer's departure time preferences, the strength of his or her departure time preferences, w, his or her willingness to pay for the product when it is offered at their preferred departure time, and whether or not he or she is fully informed, are all independently distributed. However, consumers' departure time preferences are correlated across consumers so the total demand at each departure time is unknown ex ante.

[4] The assumption that uninformed consumers choose to observe only prices and products at their preferred departure time is not without loss of generality; consumers with low waiting costs strictly prefer to search at their non-preferred departure time because the expected price is lower. However, the testable predictions of the model are unchanged if we instead assume that uninformed consumers choose which information to acquire.

time, and hence these consumers will choose to fly at the off-peak time if their waiting cost, *w,* is less than the price difference.

Below we derive the equilibrium prices in the competitive and monopoly environments. In both cases, firms offer just two prices in equilibrium. Firms offer enough capacity at the low price to meet demand at the ex post off peak departure time, and ration consumers at the ex post peak departure time. This implies that low-price capacity is utilized 100% of the time (is sold with probability one). Firms offer enough capacity at the high price to meet the demand at the ex post peak departure time, which implies that high-price capacity is utilized 50% of the time, or equivalently, is sold with probability ½.[5]

Given two prices, p_L and p_H, or equivalently, a base price, $p = p_L$, and a price premium, $\Delta = p_H - p_L$, where $p + \Delta \leq V$, the capacity offered at the low price is equal to the off-peak demand, or the number of consumers who prefer the off-peak departure plus the number of switching consumers (i.e., those who are informed, prefer the peak departure, are rationed at the low price, and have waiting costs less than Δ), or

$$Q_L(p,\Delta) = \phi_L N + (\phi_H N - Q_L(p,\Delta))\alpha F(\Delta).$$

The capacity offered at the high price is equal to the number of consumers who prefer the peak departure less the number of these consumers who are served at the low price less the number of switching consumers, or

$$Q_H(p,\Delta) = \phi_H N - Q_L(p,\Delta) - (\phi_H N - Q_L(p,\Delta))\alpha F(\Delta).$$

It follows that in the equilibrium capacities are

$$Q_L(p,\Delta) = \phi_L N + (\phi_H - \phi_L) N \frac{\alpha F(\Delta)}{1+\alpha F(\Delta)}, \tag{1}$$

and

$$Q_H(p,\Delta) = (\phi_H - \phi_L) N \frac{1-\alpha F(\Delta)}{1+\alpha F(\Delta)}. \tag{2}$$

The total industry capacity at each departure time is

[5] This behavior is optimal in a neighborhood of the optimal prices, so we make this assumption without loss of generality (see Dana, 1999a).

$$Q_L + Q_H = \left[1 - \frac{\gamma}{2} + \gamma \frac{1}{1+\alpha F(\Delta)}\right] N \tag{3}$$

where $\gamma = \phi_H - \phi_L$ is a measure of the volatility of demand for each departure time (γ is proportional to the standard deviation divided by the mean). Note that $\phi_L = 1 - \gamma / 2$.

Total sales are equal to

$$2Q_L + Q_H = 2\phi_L N \frac{1+\alpha F(\Delta)}{1+\alpha F(\Delta)} + 2(\phi_H - \phi_L) N \frac{\alpha F(\Delta)}{1+\alpha F(\Delta)} + (\phi_H - \phi_L) N \frac{1-\alpha F(\Delta)}{1+\alpha F(\Delta)} = N$$

so the capacity utilization, or load factor, is

$$\mathrm{LF}(\alpha,\Delta,\gamma) = \frac{2Q_L + Q_H}{2Q_L + 2Q_H} = \frac{1}{1 + \frac{\gamma}{2}\left[\frac{1-\alpha F(\Delta)}{1+\alpha F(\Delta)}\right]}. \tag{4}$$

It follows that $\partial \mathrm{LF}/\partial\alpha > 0$, $\partial \mathrm{LF}/\partial\Delta > 0$, $\partial \mathrm{LF}/\partial\gamma < 0$, $\partial^2 \mathrm{LF}/\partial\alpha\partial\Delta > 0$, and $\partial^2 \mathrm{LF}/\partial\alpha\partial\gamma > 0$.

3.1 Competitive Pricing

In a competitive market, the equilibrium prices are the zero-profit prices, $p_L = k + c$ and $p_H = 2k + c$, or equivalently, $p_c = k + c$ and $\Delta_c = k$ where the subscript *c* denotes competitive pricing (not to be confused with the cost *c*). Evaluating load factor and capacity at these prices, equations (3) and (4) clearly imply that in a competitive market equilibrium load factors are increasing in α, i.e., $d\mathrm{LF}_C / d\alpha > 0$, and equilibrium capacity is decreasing in α, i.e., $d(Q_L + Q_H)/d\alpha < 0$.

3.2 Monopoly Pricing

Now consider the monopolist's pricing problem. Following Dana (1999a), the monopolist offers two prices, p_L and p_H, or equivalently, p_m and $p_m + \Delta_m$, where the subscript *m* denotes monopoly pricing.

The monopolist chooses p_m and Δ_m to maximize its profits, however clearly $p_m + \Delta_m = V$, or $p_m = V - \Delta_m$, so the monopolist's problem is

$$\max_{\Delta_m} 2\phi_L Q_L (V - \Delta_m, \Delta_m)(V - \Delta_m - k - c) + \phi_H Q_H (V - \Delta_m, \Delta_m)(V - 2k - c).$$

The first order condition is

$$-2\left(\phi_L+\left(\phi_H-\phi_L\right)\frac{\alpha F\left(\Delta_m\right)}{1+\alpha F\left(\Delta_m\right)}\right)+\left(\phi_H-\phi_L\right)\frac{2\alpha f\left(\Delta_m\right)}{\left(1+\alpha F\left(\Delta_m\right)\right)^2}\left(k-\Delta_m\right)=0,$$

which uniquely defines Δ_m as a function of α. It follows from the first order condition that $\Delta_m < k = \Delta_c$ which implies $\alpha F\left(\Delta_m\right) < \alpha F\left(\Delta_c\right)$, or, all else equal, fewer customers shift their purchases from the peak to the off-peak flight in a monopoly market than in a competitive market.

As in the case of competitive markets, the monopolist's load factor rises as α rises. This follows from (4), holding Δ_m fixed, but is true more generally because the first order condition implies $d\Delta_m/d\alpha > 0$ and so $d\mathrm{LF}_m/d\alpha = \partial\mathrm{LF}/\partial\alpha + \partial\mathrm{LF}/\partial\Delta \, d\Delta_m/d\alpha > 0$.

Also, $\Delta_m < k = \Delta_c$ and $\partial^2\mathrm{LF}/\partial\alpha\partial\Delta > 0$ implies that holding Δ_m fixed, that monopolist's load factor is less sensitive to increases in information than a competitive firm. Moreover, when F is uniform, it also follows that $d\mathrm{LF}_c\left(\alpha,\Delta_c,\gamma\right)/d\alpha > d\mathrm{LF}_m\left(\alpha,\Delta_m\left(\alpha\right),\gamma\right)/d\alpha$.

3.3 Theoretical Predictions

The theory predicts that the expected load factor is decreasing in the volatility of the market, γ, increasing in the extent of product and price information, α, increasing in the number of firms in the market, *n*, increasing in the elasticity of departure time preferences with respect to the price differential, $\upsilon = \Delta f\left(\Delta\right)/F\left(\Delta\right)$, and increasing in the cost of capacity, *k*. The theory also predicts that greater competition, *n*, and volatility, γ, will increase the sensitivity of load factor to product and price information. More succinctly, the model predicts that

$$\mathrm{E}\left[\mathrm{LF}\right]=\frac{\mathrm{E}\left[\mathrm{Sales}\right]}{\mathrm{Capacity}}=F\left(\overset{-}{\gamma},\overset{+}{\alpha},\overset{+}{n},\overset{+}{k},\overset{+}{c},\overset{+}{\upsilon},\overset{+}{\gamma\alpha},\overset{+}{n\alpha}\right). \tag{5}$$

3.4 Robustness

The theoretical model is quite stylized and includes several potential restrictive assumptions. First, we considered a model in which capacity and prices were set ex ante before firms learned demand. Ex post pricing is more realistic in many contexts, but in the airline industry where many tickets are purchased in advance, we think ex ante pricing is more realistic. Most importantly, all of our theoretical predictions also hold in a model with ex post pricing. The important assumption is that firms set capacity before they know demand.

Another strong assumption is that uninformed consumers choose to search at their preferred departure time. In a more realistic model in which consumers searched optimally, a few consumers with very low waiting costs would choose to search first at their non-preferred departure time. However, all of our theoretical predictions also hold in a model with optimal search.

Another limitation was that we looked only at the monopoly and competitive models and did not actually analyze an oligopoly model. Our comparison of the monopoly model and the competitive model reveals firms with greater market power have lower load factors, but we did not consider an oligopoly model because in this simple model with just two discrete demand states, firms' equilibrium strategies would have been mixed.[6]

We also chose to focus exclusively on uncertainty about departure time preferences and ignored uncertainty about the level of demand. A more general model that included uncertainty about the level of demand would also predict that the expected load factor was increasing in the elasticity of demand; firms facing relatively fewer price inelastic customers make relatively fewer speculative investments in capacity.

In addition, the model ignores asymmetries across firms. In an oligopoly model with heterogeneous costs or product differentiation, firms will have different equilibrium market shares. While asymmetries could affect load factors in several ways, an important one is that firms with competitive advantages will draw a larger share of the market and face a less volatile demand then firms with smaller shares (we discuss the relationship between volatility and volume further in the estimation section below).

Finally, the model does not capture some other 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).

3.5 Welfare

Note that social welfare increases as α increases. First, for many consumers using the Internet is a lower cost way to search, so holding the amount of information gather fixed, consumers have lower search costs. Second, an increase in α leads to increases in consumer surplus by allowing consumers to choose among a greater variety of prices and products. And third, an increase in α leads to an increase in consumer switching, reduces the ex post peak demand, and reduces the capacity that airlines need to offer to meet demand.[7]

In the competitive model, profits are zero so the increase in social welfare is equal to the increase in consumer surplus. For every additional consumer who switches, costs fall by $2k$. These switching consumers also bear a waiting cost, w, because they switch to their non-preferred departure time. However since they switch voluntarily, it follows that $w < k$. So consumer surplus (and social welfare) increases by $2k - \mathrm{E}[w|w < k] > k$ per switcher. When w is uniformly distributed on $[0,\bar{w}]$ for $\bar{w} > k$, this implies that the social welfare gain from an increase in consumer information is ¾'s of the cost savings and regardless of the distribution is at least ½ of the cost savings.

In some respects, our theoretical model may overstate the welfare gains from increased consumer information. First, if prices are set ex post, after firms learned demand, then in the competitive model the increase in consumer surplus is $2k - \mathrm{E}[w|w < 2k]$ per consumer, which is strictly positive but nevertheless smaller than $2k - \mathrm{E}[w|w < k]$. When w is uniformly distributed on $[0,\bar{w}]$ for $\bar{w} > k$, this implies that the social welfare gain from only ½ of the cost savings, but across all distributions there exists no theoretical lower bound on the welfare gain.

---

[6] A more general model with a continuum of demand states and a pure-strategy equilibrium is beyond the scope of this paper (see Dana, 1999b).

[7] Another potential cost of an increase in consumer information is greater passenger congestion, however it may also lower airport congestion and improve on-time performance.

Second, some of the increase in capacity utilization could come from consumers with low valuations as opposed to low waiting costs. Still, when prices are set ex ante, revealed preference implies the social welfare gain is at least k per new consumer.

However, in other respects our model clearly understates the welfare gains from increased consumer information. Most importantly, we do not include the direct effect on consumers lower search costs.

Also, the Internet may have significantly lowered the costs in the airline industry. Clearly ticketing and distribution costs have fallen as online ticketing has eliminated travel agents and travel agency fees and even reduced costs relative to telephone reservation systems. This social welfare benefit is neither included in our stylized model nor measured in our empirical analysis.

4. Data

We use multiple data sources. 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 airline-airport-pair segment*,* or leg, and by aircraft type, for all domestic passenger flights in the US. A directional airline airport-pair segment includes all the flights that travel nonstop from one airport to the other (a single take-off and landing). The capacity and passenger traffic data in the T100 database are used to calculate the average load factor for each airline on each directional airport-pair segment.[8] We also use the T100 data to obtain travel distance for each airport-pair segment and to obtain an historical measure of segment size.

Our second data source is the Computer Use and Ownership Supplement to the Consumer Population Survey (CPS). We use the CPS to measure Internet penetration for every major metropolitan area. 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.

[8] Load factor is commonly defined as revenue passenger miles divided by available seat miles. On a single segment this is equivalent to revenue passengers divided by available seats, so we ignore the distinction.

Our third data source is the Origin and Destination Survey (DB1B) market database. This is a 10% sample of all passenger tickets 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 fare, the number of passengers paying the fare, the origin and destination airports (for the passenger), and the itinerary (the individual flight segments flown). The DB1B market database includes two entries for each roundtrip ticket and one entry for each one-way ticket. Importantly, the DB1B database identifies which entries are the outbound and return portions of round-trip tickets, so we know which airport is the passenger's home airport (and therefore the associated metropolitan area in which he or she is likely to have purchased his or her ticket). However, before 1999, Southwest Airlines reports all of its roundtrip ticket sales as two one-way tickets, so we cannot identify the home airport for these Southwest passengers in the DB1B database.

For simplicity we restrict the DB1B database to customer itineraries with at most one stop on each directional market. We also drop itineraries on which the carrier on any segment was unknown, itineraries with "top-coded" fares, and itineraries with fares below $25 in 2000 dollars. We also dropped itineraries with a total travel distance less than 50 miles. These restrictions allow us to limit our analysis to economically significant markets and avoid introducing noise or bias associated with data entry errors.

We use the DB1B dataset for several purposes. First, we use the DB1B dataset to calculate the metropolitan-area traffic weights used to match our metropolitan area data, in particular Internet penetration, to our segment data. Simply matching the segment data to the metropolitan area in which the flights' origination airport is located is inadequate. First, many passengers are returning home on the return portion of a round-trip ticket, so these passengers are just as likely to have purchased their ticket in the metropolitan area in which the flight's destination airport is located. Still other passengers are on the second leg of their outbound itinerary (or the first leg of their return itinerary) so the airport at which these passenger began their round-trip travel is neither the airplane's origination nor destination airport. The distinction is important because our hypothesis is that the Internet penetration in the metropolitan area where passengers live and purchase their tickets affects load factors, not the Internet penetration in the metropolitan area in which the plane originates its flight.

The metropolitan-area traffic weights constructed from the DB1B database are equal to fraction of each airline's passengers flying on each segment that originate their one-way or round trip itinerary in each metropolitan area. We use the DB1B database to find all of the passengers with itineraries that include the particular airline airport-pair segment and then calculate the fraction of these consumers whose itineraries originated at each airport. A metropolitan-area traffic weight is the sum of the airport weights across all airports located in the metropolitan area. Finally we compute the traffic-weighted average of the metropolitan area Internet penetration to obtain a measure of Internet penetration that is unique to each airline airport-pair segment.[9]

Second, we use the DB1B dataset to calculate airlines' market shares. As is standard in the literature, our market definition is a directional origin-and-destination market, which includes all passenger travel from the origin airport to the destination airport, including non-stop flights (a single segment or leg) and connecting flights (two segments or legs). The distinction between market structure within a segment and market structure within an origin-and-destination market is important because clearly an airline can be the only carrier flying on the A-to-B and B-to-C segments yet face a great deal of competition in the A-to-C market. Specifically we calculate each airline's share of each origin-and-destination market as its share of the total traffic, including all one-way itineraries from the origin airport to the destination airport, the outbound potion of all round-trip itineraries from the origin airport to the destination airport, and the inbound portion of all round-trip itineraries from the destination airport to the origin airport. We use these market shares to construct definitions of market structure (dummy variables for monopoly, duopoly, and competition) for each origin-and-destination market.

---

[9] For example, consider a flight from airport A to airport B. Assume that 40 percent of the passengers are flying round trip from A to B, so they are on the outbound portion of their round-trip itinerary; another 35 percent of the passengers are flying round trip from B to A, so they are on the return portion of their round-trip itinerary; another 15 percent of the passengers are flying round trip from airport C to airport B with a stop each way in airport A, so they are on the second segment of the outbound portion of their round-trip itinerary; and finally, the remaining 10 percent are flying round trip from airport A to airport D with a stop each way in airport B, so they are on the first segment of the outbound portion of their round-trip itinerary. Then the weighted average Internet penetration for passengers on this particular airline's flight from airport A to airport B is equal to $(0.40 + 0.10)\ IP_A + 0.35\ IP_B + 0.15\ IP_C$, where IP_i denotes the Internet penetration in the metropolitan area in which airport i is located.

Third, we use the DB1B dataset to construct the origin-and-destination-market traffic weights we use to match the origin-and-destination-market definitions of markets structure to our segment data. Specifically we consider every passenger flying on a particular directional-airline-airport-pair segment and calculate the share of those passengers that are customers in each possible origin-and-destination market.

Fourth, we calculate the average fare paid for each directional airline airport-pair segment. To do this, we first allocate the fare paid for each consumer's itinerary among the itinerary's segments in proportion to the distance flown and then average these fares across passengers who flew on that directional-airline-airport-pair segment. Note that because the fare is allocated in proportion to the distance flown, this is an imperfect measure of the incremental cost to a consumer of flying on the segment.

And finally, we use the DB1B dataset to calculate some airline characteristics. For each directional-airline segment we calculate the fraction of the passengers whose itinerary is non-stop and the total number of the airline's itineraries that include the segment. And for each airline at each airport we calculate the number of other airports served with non-stop flights by that airline from that airport and the fraction of the airline's itineraries that are direct.

Another data source is the Bureau of Economic Analysis for metropolitan area demographic and economic data used as controls for expected and unexpected demand changes that may be spuriously correlated with Internet penetration. We use the traffic weights described above to match this data to the directional-airline segments. We also obtained wholesale jet fuel monthly price series from the US Department of Energy.

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. Note again, that Southwest Airlines was excluded because of the reporting issues noted above. We also removed the 3rd and 4th quarters of 2001 from our sample because of the events on 9/11/2001, which severely disrupted airline service in the 3rd and 4th quarters of 2001. This leaves us with 85568 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 log-log version of equation (5), where the level of observation is airline *i* on segment *j* in quarter *t*,

$$\ln \mathrm{LF}_{ijt} = \beta_I \ln \mathrm{IP}_{jt} + \beta_D X^D_{jt} + \beta_C X^C_{ijt} + \beta_{MS} \mathrm{MS}_{ijt} + \beta_{IP \cdot MS} \ln \mathrm{IP}_{jt} \cdot \mathrm{MS}_{ijt} + \beta_{IP \cdot S} \ln \mathrm{IP}_{jt} \cdot \sqrt{\mathrm{S}_j} + \xi_{it} + \varsigma_{ij} + \varepsilon_{ijt}.$$

This is a reduced-form regression of airline, airport-pair capacity utilization (LF) on traffic-weighted measures of metropolitan area Internet penetration (IP) and market structure (MS); on demand (X^D) and cost (X^C) controls; on interactions between Internet penetration and market structure ($\mathrm{IP} \cdot \mathrm{MS}$); and between Internet penetration and segment size ($\mathrm{IP} \cdot \sqrt{\mathrm{S}}$); and on airline-segment (ξ_{jt}) and airline-quarter (ς_{it}) fixed effects.

We use a log-log specification 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, throughout our analysis, we weight our observations by the number of available seats. While our unit of observation is an airline, directional, airport-pair quarter, the economic unit of observation that is of interest is a seat. Particularly because our goal is to make welfare calculations, we need to put more weight on airport-pairs with more flights and more available seats.

The dependent variable in our analysis is the logarithm of the quarterly, airline, directional, airport-pair load factor. This is a measure of the average *realized* capacity utilization as opposed to the *expected* capacity utilization. Because of this difference, we are introducing additional noise that is correlated with unexpected short-run demand fluctuations. For this reason our regression includes controls for unexpected changes in demand as well as changes in expected demand. Failure to control for unexpected demand could bias our estimates if other independent variables are correlated with unexpected demand.

The main independent variable is the measure of Internet penetration, which, as discussed above, is specific to each directional airport-pair segment and each quarter (although the measure of metropolitan area Internet penetration changes annually and is the same across airports in the same metropolitan area, the traffic weights vary by airport and change quarterly). That is,

$$\mathrm{IP}_{ijt} = \sum_k \omega_{ijkt} \mathrm{I}_k^m \mathrm{IP}_{mt} \ ,$$

where ω_{ijkt} is the share of all passengers that fly on airline *i*'s flight serving segment *j* in quarter *t* who began the travel itinerary in airport *k,* IP_{mt} is the internet penetration in metropolitan area *m*, and I_k^m is an indicator that is equal to 1 if airport *k* is located in metropolitan area *m* and is equal to 0 otherwise.

The market size and demand controls are metropolitan area population, metropolitan area employment as a percentage of population, and average per capita income. Each of these measures is matched to the segment data using the same weights as used above to match Internet penetration to the segment data. These variables are included to control for both short and long run variations in demand growth across airline-segments and hence are constructed from the MSA data in the same way as Internet penetration.

Our main controls for cost shocks are the airline-quarter fixed effects, since we expect most cost changes to be common across segments, but we also include the product of fuel costs and segment length and measures of scale and scope at the origin and destination airport. The latter include the number of other airports served with non-stop flights by that airline from both the origin and destination airport, the fraction of the airline's passengers on the segment who are transferring at one of the airline's hubs, and the fraction of the airline's total passengers departing the origin (and destination) airport whose itineraries are non-stop. Note that these variables control for product characteristics as well as costs and hence may influence load factor either through their impact on costs or demand.

Another important set of independent variables is market structure and the interaction between market structure and Internet penetration. As we discussed above market power is more accurately measured at the origin-and-destination market level rather than the segment level; therefore, we construct traffic-weighted measures of market structure for each airline, airport-pair segment. Our measures of market structure are the fraction of passengers that are traveling in a monopoly origin-and-destination market, the fraction of passengers are traveling in a duopoly origin-and-destination market, and the fraction of passengers that are traveling in a competitive origin-and-destination market. More formally, the fraction of passengers traveling in a monopoly market is

$$\mathrm{M}_{ijt} = \sum_{r} \varpi_{ijrt} \mathrm{M}_{rt} ,$$

where ϖ_{ijlt} is the share of all passengers traveling on segment j on a flight offered by airline i in quarter t whose directional-origin-and-destination itinerary is part of market r; and M_{rt} is a dummy variable equal to one if market r is a monopoly market in quarter t. A monopoly market is defined as a market in which the largest firm's market share (share of passengers) exceeds 90%. The market structure variables D_{ijt} and C_{ijt} are similarly defined, where a duopoly market is a non-monopoly market in which either the two largest firms' combined market shares exceeds 90% or the two largest firms' combined market shares exceeds 80% and the third largest firm's share is less than 10%; and a competitive market is every market which is neither a monopoly market nor a duopoly market. Note that our segment-level market structure variables are airline specific because the weights are airline specific, even though the underlying market-level market structure variables are not airline specific.

The expected volatility of demand is an important determinant of load factors that we do not directly observe. However, one important measurable source of variation in demand volatility is segment size, which we exploit when we explore whether the impact of Internet penetration varies with segment size. If the aggregate demand distribution is the sum of independent binomial decisions, then 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. Because contemporaneous measures of size are endogenous, we use an historical measure of size. Our measure of segment size is a proxy variable equal to average traffic in the same quarter of the 1994 calendar year. Of course, the impact of segment size is not identified because we also include airline-segment fixed effects; however, the interaction between Internet penetration and segment size is identified.

Since we have limited data on costs, the volatility of demand, the elasticity of demand with respect to price, and the elasticity of departure time preference with respect to the price differential, 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-varying airline, metropolitan area, airport, and airport-pair segment 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, contract driven costs (such as labor, fuel, and aircraft maintenance), and capacity (such as differences in capacity planning that lead to variation in the shadow cost of capacity).

Some sources of unobserved variation are not adequately controlled for with our fixed effects. The most important of these are probably segment-specific changes in characteristics of demand, including the level, volatility, and composition of market demand. Other important sources of unobserved variation are changes in the degree of firm rivalry and the threat of entry.

5.1 Instruments

Several of our independent variables, including market structure, airline's segment share, average segment fare, as well as interactions between market structure variables and Internet penetration, are all potentially endogenous. In this subsection we discuss the instruments that we used in our regression analysis.[10]

The instruments we chose for these potentially endogenous variables are largely based on the discrete choice literature (see, for instance, Berry, Levinsohn and Pakes, 1995, or Bresnahan, Stern and Trajtenberg, 1997) and literature on the airline industry (see Borenstein, 1989, and Peters, 2006). Specifically, we use instruments for market structure that are broader measures of market size and competitiveness, including the average population of the origin and destination cities across all itineraries that include the segment, and the number of carriers who serve the segment; and characteristics of the airline's rivals' networks that affect demand, including the average fraction of rivals' passengers who are flying in monopoly, duopoly, and competitive directional-origin-and-destination markets,

[10] The weights that we use to match metropolitan area Internet penetration (and other metropolitan area variables) to the segment level observations are also potentially endogenous. While we don't know how to directly control for endogenous weights, we repeated our analysis using weights that are fixed over time at their average levels and we obtained very similar results.

the average fraction of the airline's rivals' passengers that are flying nonstop, and the average of the airline's rivals' total itineraries that include the segment.

Our instruments for average segment fare are the average segment fare on all other segments of a similar length (we divide segments by length into five quintiles), and the airline's rivals' average fare on the reverse segment.

5.2 Results

In our first set of regressions in Table 4 we ignore the issue of endogeneity and concentrate on qualitative relationship between the variables. All of our regressions include airline-quarter and airline-segment fixed effects. The regression in Column 1 includes only Internet penetration and finds a large and statistically significant effect of Internet penetration on load factors. The coefficient of 0.08 implies that each percentage point increase in Internet penetration increases load factors by 0.08%.

In Column 2 we introduce our basic controls for expected and unexpected demand, and for capacity costs. This is a long run, reduced-form regression and measures the long-run equilibrium impact of the Internet under the assumption that market structure, capacity, and prices are all exogenous. The coefficients of our controls are consistent with our predictions. Higher fuel costs increase the cost per available seat and increase load factors and demand shocks have a positive effect on load factors.[11]

When we include our controls, the coefficient on Internet penetration is 0.107, which is even larger then in Column 1. Recall that our dependent variable is realized load factor and not expected load factor, so a potential source of bias in our regression is positive correlation between unobserved, unanticipated demand and Internet penetration. This is because unanticipated demand shocks are clearly correlated with average quarterly load factors. And this is one reason why we include traffic-weighted population at the passengers' origin city and traffic-weighted average income as controls for unanticipated demand. However, we find that including these controls actually increases the coefficient on Internet penetration, which suggests that the coefficient on Internet penetration is probably

[11] Variation in household income is likely to be inversely correlated with the elasticity of consumers' departure time preferences, and negatively correlated with load factor but we think most of this variation is constant over time and being captured by the fixed effects.

not being significantly biased by our failure to control more completely for unanticipated demand shocks.

In Column 3 we estimate a medium run, reduced-form regression including the fraction of passengers on a segment whose travel is in monopoly and duopoly markets as additional controls. While market structure is endogenous, it is likely that much of the variation in market structure over time will be driven by the financial conditions of the airlines and by demand changes in other markets. As predicted by the theory, we find that load factors are lower on segments that serve more concentrated markets. A segment that serves passengers flying in monopoly markets is estimated to have a 7.4% lower load factor than a segment that serves passengers flying in competitive markets, and a segment that serves passengers flying in duopoly markets is estimated to have a 3.0% lower load factor than a segment that serves passengers flying in competitive markets. Both of these differences are statistically significant. Intuitively, firms with greater market power have higher margins and therefore greater incentive to hold speculative capacity.

In Column 4 we include Internet penetration interacted with the fraction of passengers on a segment whose travel is in monopoly and duopoly markets. In these regressions the coefficient on Internet penetration represents the impact of Internet penetration in competitive markets. The coefficients on the interaction terms are all statistically significant and also statistically different. The impact of Internet penetration increases with competition as predicted by the model; the effect of Internet penetration is largest on segments on which all of the passengers are traveling in competitive directional-origin-and-destination markets, followed by segments on which all of the passengers are traveling in duopoly directional-origin-and-destination markets, and then followed by segments on which all of the passengers are traveling in monopoly directional-origin-and-destination markets. Again, the differences in the effects across different market structure are statistically significant. Intuitively, changes in Internet penetration have a greater impact on load factors in more competitive markets because firms with market power offer fewer incentives to consumers to switch departure times (they wish to avoid cutting prices to non-switchers in order to induce switching) so an increase in informed consumers leads to less switching.

As discussed earlier, our only measure of market volatility is historical segment size. Nevertheless, since the theory predicts that the impact of Internet penetration on load factors is higher in more volatile markets, we can exploit the time series and cross-sectional variation in Internet penetration to estimate the coefficients on the interaction between Internet penetration and the square root of market size. In Column 5, we include the square root of segment size times the log of Internet penetration as an additional dependent variable. Consistent with our theoretical model, the coefficient on this interaction is positive and statistically significant. That is, the impact of Internet penetration is larger in lower volume segments.

In Table 5 we estimate the last three regressions from Table 4, but this time using our instruments for the market structure variables. The instrumental variables estimation has little effect on our estimate of the impact of the Internet. The coefficient in the regressions without the market structure interactions is unchanged, equal to 0.106, and in the regressions with the market structure interactions it falls slightly from 0.121 to 0.119. Note though that the coefficients on market structure are somewhat larger in the instrumental variables regressions while the coefficients on the interactions between market structure and Internet penetration are somewhat smaller.

5.2 Average Fare and Segment Share

In Table 6, we include two additional controls that are not part of our theoretical analysis: average fare and segment share. While these variables are endogenous, they are correlated with important unobserved segment characteristics. Average fare is inversely correlated with unobserved heterogeneity in demand elasticity and unobserved heterogeneity in firm rivalry, while segment share is correlated with differences in firm size and unobserved differences in cost and demand between airlines.

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 the biggest source of variation in fare levels is likely to be costs, these results are not surprising since we control for costs with airline-segment fixed effects, so the remaining variation in fares over time is likely to be because of segment-specific changes in the elasticity of demand, in the intensity of firm rivalry, or in the threat of entry rather than segment-specific changes in cost.

Controlling for average fare also helps us better understand the mechanism by which Internet penetration is affecting load factors. Our theory is that giving consumers additional information increases their consumption of flights at ex post off-peak departures times, which in turn increases capacity utilization. However a plausible alternative is that price information increases price competition, which in turn reduces the incentives for airlines to invest in capacity.

The regression results in Table 6 suggest that both explanations are true. In Column 1, adding average fare as a control variable reduces the coefficient on the Internet variable from .107 to .075. That is, holding fare fixed, Internet penetration has a positive and significant effect on load factors. So the Internet appears to be diverting traffic holding average fare fixed. However the total impact of the Internet is much larger, suggesting that the increases in Internet penetration also lead to higher load factors through increased price competition. However, it is also possible that average fare falls when Internet penetration rises because average costs are falling and not because increased price competition.

In Column 2, we find that load factors are higher for firms with a larger segment share within each segment. While our model ignored differences between firms, the data suggests these differences are very important. This is consistent with our intuition about the impact of segment size; firms with larger segment share may face less volatile demand and find it easier to match capacity to demand. That is, it is easier to match the number of planes to the number of passengers when a firm has more passengers. On the other hand, firms with a larger segment share may also have a superior product or lower costs. However, a firm with lower costs, or a superior product, is likely to have higher margins and lower load factors, which is inconsistent with our estimates. However, clearly much more work can be done to better understand the role of firm asymmetries in predicting capacity utilization.

In Columns 3 and 4, we introduce both the average fare on a segment and our market structure variables. The results do not change substantially from previous specifications. While these variables are all endogenous, they are likely to be correlated with other unobserved exogenous variables. The fact that our estimate of the impact of Internet penetration is 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 unobserved market characteristics.

Finally, in Column 5 we estimate a regression with the average fare on a segment and our market structure variables using our full set instruments for market structure and fare and find qualitative similar results. Not surprisingly, the coefficient on average fare is significantly larger.

6. Conclusion

In all of our regressions analysis, we find that Internet penetration has a positive and statistically significant effect on load factors. Using the results in Column 2 of Table 4, the elasticity of Internet penetration on load factor is 0.107. That is, each percentage point increase in Internet penetration increases load factors by .107%, and a doubling in Internet penetration increases load factors by 7.7% (i.e., $2^{.107} - 1$). 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 airlines' load factors during our sample period.

Total US airline industry passenger flying operations and maintenance costs were $40 billion in 2000, so an increase of 7.7% in load factor represents approximately a 7.7% decrease in these costs, or over $3 billion in cost savings every year.

The Internet has 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 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, and probably also a much larger set of options than those that are available to a customer making a reservation through a travel agent. 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. Undoubtedly United

Airlines is able to capture some of the surplus created when it induces consumers to switch flights, so it is interesting to note that it is United, not Expedia, which offers this feature.

Because consumers are more informed, consumers are more likely to take advantage of inducements to fly at off-peak times, particularly in more competitive markets, in which firms are more likely to offer incentives for switching because the efficiency gains are more likely to outweigh the lost rents from passengers who choose to fly off-peak without incentives. Similarly, Internet penetration has a larger effect on smaller segments, in which firms are more likely to offer incentives for switching because they have a more difficult time forecasting demand. Consistent with this theoretical prediction, we find that differences in Internet penetration across time and metropolitan areas are positively correlated with differences in load factors, and that the magnitude of the relationship between Internet penetration and load factor is greater in segments that serve more competitive markets and greater in segments that have less traffic historically.

While increases in Internet access have lead to increases in airlines' load factors and a decrease in airlines' costs of over $3 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, 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 argue that half or more of this costs 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.

Berry, S., Levinsohn, J. and Pakes A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 60(4), pp. 889–917.

Bresnahan T.M., Stern S. and Trajtenberg M. (1997), "Market Segmentation, Transitory Market Power, and Rents from Innovation: Personal Computers in the late 1980's", *RAND Journal of Economics*, 28 (0), pp. S17-S44.

Borenstein S. (1989), "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry", *RAND Journal of Economics*, 20(3), pp. 344-365.

Brown, G., Jr. and Johnson, M. B. (1969), "Public Utility Pricing and Output Under Risk," *American Economic Review,* 59, 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.

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, pp. 521-540.

Carlton, D. W. (1977), "Peak-Load Pricing with Stochastic Demand," *American Economic Review,* 67, pp. 1006–1010.

Dana Jr., J. D. (1998), "Advanced-Purchase Discounts and Price Discrimination in Competitive Markets," *Journal of Political Economy,* 106, 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, 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, pp. 632–660.

Deneckere, R, Marvel, H. and Peck, J. (1997), “Demand Uncertainty and Price Maintenance: Markdown as Destructive Competition,” *American Economic Review*, 87, pp. 619–641.

Ellison, G. and Ellison, S.F. (2005), “Lessons about Markets from the Internet,” *Journal of Economic Perspectives,* 19, 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,” working paper.

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.

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.

Peters C. (2006), “Evaluating the Performance of Merger Simulation: Evidence from the U.S. Airline Industry,” *Journal of Law and Economics*, 49(2), pp. 627-649.

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, 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.

Prescott, E. C. (1975), "Efficiency of the Natural Rate," *Journal of Political Economy,* 86, pp. 1229–1236.

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

	Average Segment Fare	Average Load Factor	Average Segment Share
Air Wisconsin Airlines	182.78	0.681	0.442
AirTran	90.51	0.675	0.262
Alaska Airlines	129.81	0.677	0.629
America West	111.61	0.689	0.492
American Airlines	183.60	0.692	0.625
American Eagle	161.34	0.646	0.444
ATA Airlines	124.44	0.727	0.752
Atlantic Southeast Airlines	163.57	0.677	0.168
Comair	152.40	0.665	0.516
Continental Airlines	178.67	0.713	0.789
Delta Airlines	148.24	0.702	0.725
Frontier Airlines	132.94	0.613	0.214
JetBlue	120.74	0.823	0.738
Mesaba Airlines	194.40	0.569	0.309
Northwest	167.98	0.692	0.822
Spirit Airlines	113.94	0.767	0.263
Trans World Airlines	139.73	0.683	0.760
United Airlines	185.01	0.702	0.633
US Airways	151.22	0.670	0.835

Notes: Each cell contains (weighted by number of seats) average values over each airline's directional segments in the sample over 18 quarters. Fares are in 2000 dollars. Airlines' segment shares are defined as their share of the total number of paying passengers on the segment.

Table 3. Descriptive Statistics (85568 observations)

Variable	Mean	Std. Dev.	Min	Max
LOAD FACTOR	0.674	0.161	0.003	1.000
INTERNET	0.536	0.170	0.053	0.854
Demand Variables:				
EMPLOYMENT (%)	0.605	0.037	0.373	0.775
LOG (INCOME PER CAPITA)	10.385	0.100	9.528	10.897
LOG (POPULATION)	15.090	0.647	11.454	16.744
Cost Variables:				
FUEL * DISTANCE	65.287	50.738	2.315	512.419
LOG (# CITIES NON-STOP, ORIGIN)	39.122	32.658	1.000	140.000
LOG (# CITIES NON-STOP, DEST)	38.898	32.544	1.000	140.000
% PASS TRANSFER AT HUB	0.359	0.318	0.000	1.000
% DIRECT ITINERARIES, ORIGIN	0.115	0.096	0.006	1.000
% DIRECT ITINERARIES, DEST	0.116	0.096	0.006	1.000
Market Structure Variables:				
% PASS. IN MONOP. MARKETS	0.077	0.163	0.000	1.000
% PASS. IN DUOP. MARKETS	0.443	0.323	0.000	1.000
% PASS. IN COMP. MARKETS	0.480	0.327	0.000	1.000
Additional Variables:				
SEGMENT SIZE	99641.880	96991.930	30.000	739529.000
SEGMENT FARE ($)	164.203	79.532	3.714	1718.000
SEGMENT SHARE	0.585	0.369	0.000	1.000

Table 4. OLS Regression Results

Dependent Variable:	LOG (Load Factor)				
	(1)	(2)	(3)	(4)	(5)
LOG (INTERNET)	.080***	.107***	.106***	.121***	.128***
	(.018)	(.021)	(.021)	(.021)	(.023)
EMPLOYMENT (%)		.414**	.435**	.492***	.446**
		(.178)	(.178)	(.175)	(.184)
LOG (INCOME PER CAPITA)		.589***	.570***	.563***	.575***
		(.064)	(.063)	(.062)	(.068)
LOG (POPULATION)		.006	.008	.011	.009
		(.014)	(.014)	(.014)	(.015)
FUEL * DISTANCE		.026***	.027***	.020***	.024***
		(.007)	(.007)	(.007)	(.007)
LOG (# CITIES NON-STOP, ORIGIN)		.027***	.028***	.028***	.026***
		(.006)	(.006)	(.006)	(.006)
LOG (# CITIES NON-STOP, DEST)		.023***	.024***	.023***	.021***
		(.005)	(.005)	(.005)	(.005)
% PASS TRANSFER AT HUB		.019	.008	.009	.003
		(.027)	(.028)	(.028)	(.029)
% DIRECT ITINERARIES, ORIGIN		-.084	-.093*	-.087*	-.07
		(.052)	(.051)	(.051)	(.053)
% DIRECT ITINERARIES, DEST		-.067	-.076	-.070	-.047
		(.047)	(.046)	(.046)	(.048)
% PASS. IN MONOP. MKTS.			-.040***	-.093***	-.041***
			(.010)	(.014)	(.010)
% PASS. IN DUOP. MKTS.			-.024***	-.043***	-.025***
			(.005)	(.007)	(.005)
LOG (INTERNET) * % PASS. IN MONOP. MKTS.				-.074***	
				(.014)	
LOG (INTERNET) * % PASS. IN DUOP. MKTS.				-.030***	
				(.007)	
LOG (INTERNET) * SQRT (SEGMENT SIZE)					-.034**
					(.017)
Observations	85568	85568	85568	85568	78870

Notes: Standard errors are in parentheses. Stars denote the significance level of coefficients: *** - 1 percent, ** - 5 percent, * - 10 percent. All reported regressions include airline-segment and airline-quarter fixed effects. The regression in column (5) contains fewer observations than the other regressions because segment size is not defined for all segments.

Table 5. IV Regression Results

Dependent Variable:	LOG (Load Factor)		
	(1)	(2)	(3)
LOG (INTERNET)	.106***	.119***	.128***
	(.021)	(.021)	(.023)
EMPLOYMENT (%)	.414**	.456***	.425**
	(.178)	(.177)	(.184)
LOG (INCOME PER CAPITA)	.576***	.572***	.580***
	(.063)	(.062)	(.067)
LOG (POPULATION)	.007	.009	.007
	(.014)	(.014)	(.015)
FUEL * DISTANCE	.028***	.023***	.025***
	(.007)	(.007)	(.007)
LOG (# CITIES NON-STOP, ORIGIN)	.028***	.028***	.026***
	(.006)	(.006)	(.006)
LOG (# CITIES NON-STOP, DEST)	.024***	.024***	.021***
	(.005)	(.005)	(.005)
% PASS TRANSFER AT HUB	.009	.01	.004
	(.028)	(.028)	(.029)
% DIRECT ITINERARIES, ORIGIN	-.097*	-.092*	-.076
	(.051)	(.051)	(.053)
% DIRECT ITINERARIES, DEST	-.080*	-.075	-.051
	(.046)	(.046)	(.048)
% PASS. IN MONOP. MARKETS	-.074***	-.104***	-.077***
	(.018)	(.024)	(.018)
% PASS. IN DUOP. MARKETS	-.015***	-.033***	-.017***
	(.006)	(.009)	(.006)
LOG (INTERNET) * % PASS. IN MONOP. MARKETS		-.042*	
		(.022)	
LOG (INTERNET) * % PASS. IN DUOP. MARKETS		-.027***	
		(.009)	
LOG (INTERNET) * SQRT (SEGMENT SIZE)			-.034**
			(.017)
Observations	85568	85568	78870

Notes: Standard errors are in parentheses. Stars denote the significance level of coefficients: *** - 1 percent, ** - 5 percent, * - 10 percent. All reported regressions include airline-segment and airline-quarter fixed effects. The regression in column (3) contains fewer observations than the other regressions because segment size is not defined for all segments.

Table 6. Additional Regression Results

Dependent Variable:	LOG (Load Factor)				
	OLS				IV
	(1)	(2)	(3)	(4)	(5)
LOG (INTERNET)	.075***	.104***	.072***	.086***	.070***
	(.018)	(.021)	(.018)	(.019)	(.020)
EMPLOYMENT (%)	.384**	.416**	.378**	.427**	.396**
	(.174)	(.177)	(.172)	(.169)	(.173)
LOG (INCOME PER CAPITA)	.677***	.563***	.659***	.652***	.701***
	(.063)	(.064)	(.063)	(.063)	(.066)
LOG (POPULATION)	-.005	.009	-.003	.000	-.007
	(.015)	(.014)	(.015)	(.015)	(.016)
FUEL * DISTANCE	.007	.024***	.003	-.002	-.009
	(.006)	(.007)	(.006)	(.006)	(.007)
LOG (# CITIES NON-STOP, ORIGIN)	.027***	.024***	.022***	.022***	.023***
	(.006)	(.005)	(.005)	(.005)	(.005)
LOG (# CITIES NON-STOP, DEST)	.021***	.020***	.017***	.016***	.016***
	(.005)	(.005)	(.005)	(.005)	(.005)
% PASS TRANSFER AT HUB	.029	-.003	.010	.011	.020
	(.023)	(.027)	(.022)	(.022)	(.021)
% DIRECT ITINERARIES, ORIGIN	-.170***	-.079	-.158***	-.152***	-.193***
	(.052)	(.049)	(.050)	(.050)	(.054)
% DIRECT ITINERARIES, DEST	-.141***	-.063	-.129***	-.124***	-.159***
	(.047)	(.046)	(.046)	(.046)	(.048)
% PASS. IN MONOP. MARKETS		-.065***	-.035***	-.075***	-.063**
		(.010)	(.009)	(.011)	(.029)
% PASS. IN DUOP. MARKETS		-.030***	-.022***	-.040***	-.029***
		(.005)	(.004)	(.006)	(.008)
LOG (INTERNET) * % PASS. IN MONOP. MKTS.				-.056***	-.040**
				(.012)	(.018)
LOG (INTERNET) * % PASS. IN DUOP. MKTS.				-.028***	-.026***
				(.007)	(.008)
SEGMENT SHARE		.108***	.127***	.127***	.111***
		(.015)	(.014)	(.014)	(.030)
LOG (SEGMENT FARE)	-.234***		-.238***	-.237***	-.342***
	(.012)		(.012)	(.012)	(.025)
Observations	85568	85568	85568	85568	85568

Notes: Standard errors are in parentheses. Stars denote the significance level of coefficients: *** - 1 percent, ** - 5 percent, * - 10 percent. All reported regressions include airline-segment and airline-quarter fixed effects.