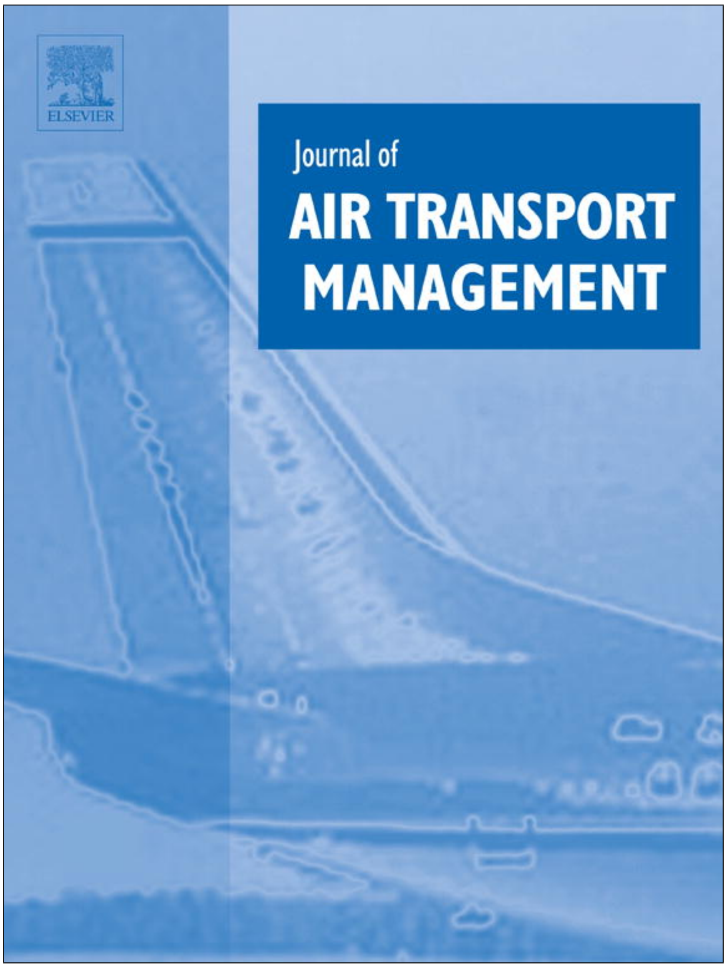
ELSEVIER
Journal of
AIR TRANSPORT
MANAGEMENT

Journal of Air Transport Management 24 (2012) 7–11

Contents lists available at SciVerse ScienceDirect

Journal of Air Transport Management

journal homepage: www.elsevier.com/locate/jairtraman

Capacity-contingent pricing and competition in the airline industry

Robert Clark*, Nicolas Vincent

HEC Montréal and CIRPÉE, 3000 Cote Sainte Catherine, Montreal, Canada H3T2A7

Keywords:
Airfare setting
Aircraft load factors
Yield-management

A B S T R A C T

We present evidence on the dynamic pricing strategies of airlines. We collect high-frequency price and load-factor data directly from the websites of the main players for one of the busiest flight corridors in the US. Using this information we characterize the determinants of pricing decisions. Our focus is on the role that competition plays in these decisions. We show that prices for some airlines are responsive to rival prices and load factors. Specifically, we find that even after controlling for the number of days remaining until the flight and the number of seats remaining, some airlines increase their prices as their rivals' available seats disappear.

1. Introduction

This paper considers the way in which firms set prices for fixed stocks of perishable products. This is a common situation experienced by airlines that must decide on prices as the date of the flight approaches and the number of available seats declines. In these situations airlines start with a fixed number of seats that will "perish" at a given and known time in the future – the date of the flight. At each point in time, they set a list price in an effort to maximize revenue. In setting price they face conflicting incentives. As their stock of seats decreases, the incentive is to increase the price to take advantage of late buyers with high willingness-to-pay. On the other hand, there is incentive to lower prices as the probability that unsold seats will be purchased in the future decreases as the flight date nears. The tradeoff depends mainly on how demand changes over time. Consumers differ in their arrival times and their willingness-to-pay for seats. When deciding whether to buy now they must forecast the evolution of seat prices until the flight date. There is a chance that they will have the opportunity to purchase the seat at a lower price. The seat may also, however, become more expensive, or may no longer be available.

This situation is further complicated by competition. With multiple airlines, consumers face numerous options when shopping for seats. Firms must consider the prices of their rivals and also their remaining stocks of seats. The competitive pressure may change over time as the stocks of some firms become depleted.

To shed light on these issues we construct a dataset of high-frequency prices and real-time capacity data using information collected directly from websites of major airlines. Our analysis is empirical, and we focus on the evolution of prices and price dispersion as the flight date approaches, of load factors and their dispersion, of prices as the realized/expected load factor changes, and of prices as the prices or load factors of rival firms change.

2. Data

Our full dataset contains information on airfares and seat availability for a number of airlines, routes, times and dates of departure. Information is collected directly from the websites of the main airlines in the US: American Airlines, Delta, United, Southwest, Continental and US Airways. A flight corresponds to an "airline/route/day and time of departure" combination and focuses exclusively on one-way flights to minimize the number of possible combinations.[1]

First, we collect price information. When prices for multiple seat types are available, such as for Southwest, the lowest price is extracted. All fares include applicable taxes. In addition to fare quotes, we also derive a measure of load factor by extracting information from seat maps for each flight.[2] Our procedure computes the number of economy seats indicated as unavailable and divides this number by the number of economy seats on the plane. While some types of seats are invariably indicated as unavailable (e.g. emergency exit row) on each flight, this systematic measurement error is absorbed by the constant term in the

* Corresponding author. Tel.: +1 5143407034.
E-mail address: robert.clark@hec.ca (R. Clark).

[1] The routes used for this analysis do not have minimum-stay or Saturday-stayover rules. In other words, one-way fares are on average half the price of return trips.

[2] Load factor information is not available for Southwest.

0969-6997/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jairtraman.2012.04.005

analysis. One might be concerned that certain blocks of seats might be withheld and then appear suddenly on the site as available. We observe no such patterns in our dataset. Our seat-map measure is obviously inferior to having daily load factor information directly from the airlines themselves, but such data are, to our knowledge, not available.

Our dataset covers flights departing on Tuesdays and Fridays on multiple dates between June 4, 2010 and March 25, 2011. For a given flight i with departure on date t, information is extracted multiple times in the months prior to t. The frequency of data extraction varies across routes. However, for the routes in the New York–Chicago corridor, data extraction was executed at a very high frequency, up to bi-hourly. We make use of this dimension of our dataset to analyze empirically very fine reaction patterns in airlines' pricing behavior.

Given the size of the dataset and to maximize data quality, we narrow our analysis to Friday flights on four routes/airlines for which we believe competition is particularly fierce: LaGuardia-O'Hare on American (AA), United (UA) and Delta (DL), as well as LaGuardia-Midway on Southwest (SW). These are very busy routes: AA, UA, DL and SW typically operated 17, 16, 5 and 15 daily flights respectively over the period.[3] The result is 695,947 observations for competing flights on which to draw relevant and reliable conclusions.[4]

In the literature, when this sort of data is used, a flight occurring at some point in the future is selected and fare observations at different points in time prior to departure are recorded from an online booking service or the airlines' websites. Compiling these data yields what is known as a "temporal-fares-offered curve." The difference here is that we have information on load factors and so can consider the evolution of capacity and of prices as both own- and rival-load factors change. Puller et al. (2009) also use a census of all transactions conducted through one of the major computer reservation systems for a third of domestic US ticket transactions and use information on fare, origin and destination, airline, flight number, dates of purchase, departure and return, booking class. A proxy of the realized load factors based on the number of tickets sold for each route and the share for their reservation system was also included, as well was an estimate of the average load factor at each point in time. Escobari (2007) uses a panel dataset obtained from Expedia.com and was able to calculate the sold out probabilities for flights; and from the T-100 from the US Bureau of Transportation Statistics he estimates average load factors at departure.

In contrast to Escobari we collect information on a large number of flight dates, which allows us to calculate expected load factors at different points prior to departure. In addition, the availability of multiple flight dates in our dataset allows us to include airline/time of day fixed effects in our regressions. Although Puller et al. do measure expected load factor, their measure is based on the assumption that on all flights the share of seats occupied by their computer reservation system is equivalent to its share in the aggregate data. More importantly, the main difference between our dataset and the one in Puller et al. is that we observe posted prices while they observe transaction prices. While there are many important advantages to using transaction data, for our purposes list prices are preferable. Since our focus is on the competitive effects that arise in dynamic pricing, we are interested in the way prices respond not only to own load factors, but also to rival-load factors and prices.

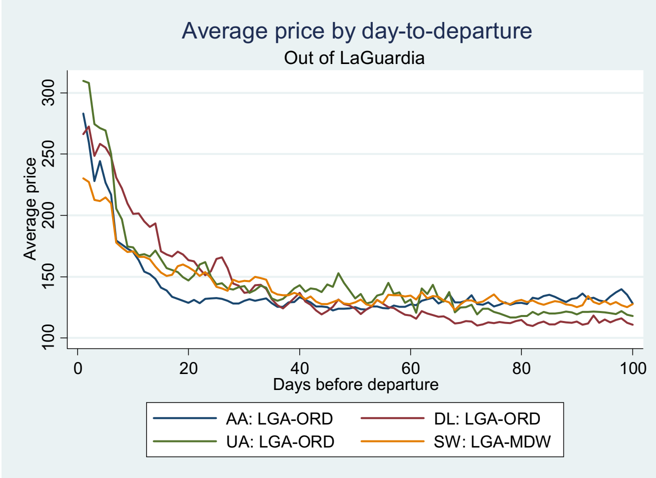

Fig. 1. Average airfare as a function of the number of days before departure.

The other contribution of our dataset and analysis is that they highlight the role of competition in the evolution of fares. There has been only limited analysis of the role of rival prices in airlines' price setting decisions (Button and Vega, 2007), and on the influence of rival-load factors. For instance, McAfee and te Velde's (2006) study of the extent to which fares of substitutes are correlated focuses only on substitute airports and substitute time slots rather than substitute carriers. They find some correlation between fares for flights at rival airports and between fares for flights on the same route but at different times. In contrast, we look specifically at whether changes in rival prices or load factors increase the likelihood of own-price changes.

3. Dynamics of prices and load factors

3.1. Prices

Fig. 1 plots the average price across departure times and dates for a given route/airline as a function of the number of days left before departure. First, we notice that ticket prices are mostly flat up to about six to seven weeks before the flight date. Then, prices tend to rise, at first modestly before strongly trending upward in the last three weeks. The result is that the average fare increases approximately twofold over the last six weeks.[5] While the general pricing pattern is comparable across airlines, there are differences. For example, fares on American Airlines remain flat for longer: the upward trend becomes noticeable only in the last 20 days. On the other hand, Delta seems to implement more gradual price increases.[6]

Additionally, prices become more variable as the departure date approaches, and this variability differs across airlines. This pattern could arise because airlines change prices more often as the flight date approaches, or because each price change is relatively larger. To distinguish between these explanations, we exploit the high-frequency nature of the data and compute the number of hours between two price changes for each route, and take the average

[3] If an airline decides to add a flight at some point, it automatically enters our dataset.

[4] This market is also chosen since both New York and Chicago are served by airports with direct flights to a very large number of destinations, the issue of connecting flights is less likely to be problematic. This could be a problem as we cannot determine whether a seat is sold as part of a multi-stop flight, in which case the observed posted price may not be relevant.

[5] This finding is consistent with results found in the empirical literature on dynamic pricing in the airline industry (Button and Vega, 2007).

[6] We do not find meaningful differences in the behavior of fares as a function of the time of departure (e.g. early morning versus middle of the afternoon).

Table 1
Duration of price spells (hours).

	Mean	Median	Standard deviation
Less than 1 week before departure			
AA	15.6	7.1	20.4
DL	24.2	12.8	27.3
SW	16.3	5.1	23.4
UA	30.6	22.1	29.3
All	19.5	9.4	24.3
1–4 weeks before departure			
AA	24.0	10.1	40.3
DL	41.4	19.3	57.7
SW	43.0	20.4	58.9
UA	55.9	28.7	65.4
All	34.9	14.2	52.5
More than 4 weeks before departure			
AA	72.9	44.6	97.1
DL	88.6	46.2	113.5
SW	113.9	56.9	152.9
UA	118.5	85.6	139.8
All	91.2	48.8	120.0

Notes: AA, DL, UA and SW correspond to American Airlines, Delta Airlines, United Airlines and Southwest respectively. Flights are from LaGuardia (LGA) to either O'Hare (ORD) or Midway (MDW).

across all flights for a given airline. This provides the average duration of a price spell: if an airline resets fares more often, the average spell will be of shorter duration.[7] Table 1 offers statistics on price spells across airlines for: the week leading to the departure date, as well as one-to-four weeks and more than four weeks before departure. First, for each airline, price spells become shorter as the flight date approaches. This fine-tuning strategy is probably implemented to maximize revenues per seat while making sure it is not left with unfilled capacity the day of departure. Also, there is important heterogeneity across airlines. For example, in the last week before departure, a typical price spell lasts only 15.6 days for AA, compared to 30.6 for UA. These differences are also present at longer horizons.

3.2. Load factors

Most prior work only analyzes the load factor at the time of departure with low frequency data (e.g. how full on average is a particular plane on a given flight over the first quarter of 2010). In contrast, our dataset allows us to track the history of load factors over time for a given flight. For example, for the 11:15am American Airline flight between LGA and ORD on September 14, 2010, we have information on the number of empty seats on the plane d days before departure in addition to the number of seats on the plane. This allows us to compute a proxy of the load factor.

Fig. 2 shows the evolution of the average normalized load factor (across flights and departure dates) for each of the three routes we focus on.[8] Not surprisingly, the fraction of seats sold generally increases over time. American Airlines and United Airlines exhibit very similar patterns, with an acceleration of the upward trend starting around 6 weeks before the departure date. Delta, on the other hand, sees a more gradual rise, starting about 80 days before departure.

When looking at the dispersion of load factors, we notice some significant differences across airlines. Fig. 3 shows how much variability there is in load factors for a given flight (e.g. 7am AA flight between LGA–ORD) across flight dates, per number of days before departure. For all airlines, dispersion tends to rise initially, between 75 and 20–30 days before departure. That is, there seems to be more uncertainty about load factors. Then, as the departure date approaches, load factors seem less and less dispersed across weeks, possibly as airlines adjust prices often to make sure the planes will be full. The fall is particularly pronounced for AA, and somewhat less so for UA.

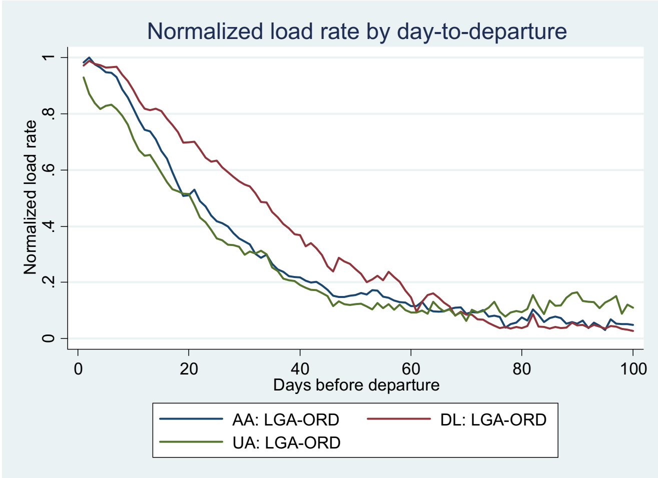

Fig. 2. Average normalized load factors as a function of the number of days before departure.

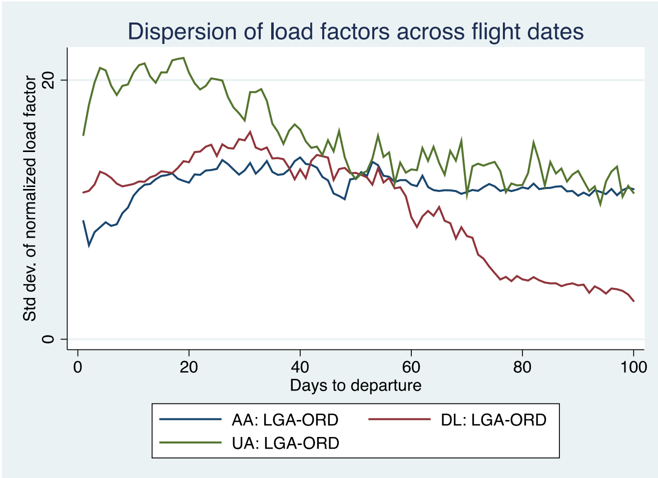

Fig. 3. Cross-sectional dispersion of load factors as a function of the number of days before departure.

4. Analysis

Table 2 reports results for basic regressions of airfares on load factors and the number of days to departure.[9] We include quadratic terms as well to capture nonlinearities. One objective is to determine whether the proxy for the load factor contains valuable information or only represents noise. In addition, we want to establish the marginal contribution of the load factor measure, since from Fig. 2 we know that it is highly correlated to

[7] Nakamura and Steinsson (2008) offer a discussion of price spells in the context of micro CPI data.

[8] Based on seat maps, load factors are never equal to zero because some seats (e.g. emergency rows) are always indicated as not available. Hence, we normalize load factors and express them on a scale of 0%–100%.

[9] One observation here represents a flight (airline/route/day and time of departure) for a given data extraction date and time.

Table 2
Determinants of airfares. Dependent variable is fares. Regressors include load factor (LOAD), number of days before departure (DAYS), along with their quadratics, as well as flight and extraction-time dummies. Robust standard errors are calculated. *, ** and *** indicate statistical significance at the 10, 5 and 1% level respectively.

	P_t^{AA}	P_t^{AA}	P_t^{DL}	P_t^{UA}
$LOAD_t$		−6.454***	−1.888***	−2.973***
$LOAD_t^2$		0.057***	0.0331*	0.0389***
$DAYS_t$	−3.808***	−0.794***	−1.6598***	−1.853***
$DAYS_t^2$	0.035***	0.008***	0.014***	0.014***
R^2	0.23	0.35	0.54	0.34

Table 3
Determinants of airfares. Dependent variable is the deviation of ticket price from usual price (*PDEV*) for this flight/days to departure. Regressors include lags of PDEV, lags of deviation of load factor from usual load factor (*LOADDEV*) and date of flight dummies. Robust standard errors are calculated. *, ** and *** indicate statistical significance at the 10, 5 and 1% level respectively.

	$PDEV_t^{AA}$	$PDEV_t^{DL}$	$PDEV_t^{sw}$	$PDEV_t^{UA}$
$PDEV_{t-1}^{AA}$	0.689***	0.071**	−0.012	0.058*
$PDEV_{t-1}^{DL}$	0.039*	0.778***	−0.011	0.035
$PDEV_{t-1}^{SW}$	−0.036	−0.036*	0.825***	−0.070**
$PDEV_{t-1}^{UA}$	0.032*	0.016	0.009	0.772***
$LOADDEV_{t-1}^{AA}$	0.003	0.073	−0.024	0.233**
$LOADDEV_{t-1}^{DL}$	0.137**	−0.092*	−0.024	0.131**
$LOADDEV_{t-1}^{UA}$	−0.013	0.107**	0.131***	−0.168**
R^2	0.86	0.88	0.95	0.89

the number of days to departure. Yet, a comparison of the first two columns shows that the addition of the load factor leads to a large increase in fit.[10] Also, load factors are strongly significant for all airlines; for instance, for a Delta flight ten days before the departure date, an increase in load factor from 50% to 60% implies an average increase in fares of about $18.

We investigate the role of competition in shaping the decision to set fares. We are interested in answering the following questions: "Is airline *i* likely to charge a price higher than usual when airline *j* already does so?" and "Is airline *i* more likely to charge higher ticket prices when its competitors have planes fuller than usual?" Our dataset is particularly well suited to study such issues because we observe high-frequency posted prices for a large number of flight dates and the universe of competitors on the LGA–CHI route. First, however, we define a range of competing flights across airlines; it would be inappropriate to consider, for example, the response of AA on its 7 am flight to the price charged by Southwest on its 9 pm flight. Hence, for the present analysis we focus on the set of flights departing between 6:30 am and 8:30 am on Friday.

For a given airline/route/flight time, we have numerous observations on fares and loads for multiple departure dates. This allows us to compute expected load factors ($LOAD^{EXP}$) and normal prices (P^{EXP}) for a given number of days before departure by averaging load factors and fares across flight dates.[11] For example, we can determine how full the LGA–ORD 7 am flight on AA normally is one week before departure, and the normal fare charged. Then, for each observation in our dataset we define $LOAD^{DEV} = LOAD - LOAD^{EXP}$ and $P^{DEV} = 100(P - P^{EXP})/P^{EXP}$ which represent the deviation of the load factor and ticket price from their usual levels for this number of days before departure.[12]

Table 4
Competition and price adjustments. Probability of a price adjustment as a function of deviation from usual load rate (*LOADDEV*), competitors' past price adjustments and days-to-departure dummies. Robust standard errors are calculated. *, ** and *** indicate statistical significance at the 10, 5 and 1% level respectively.

	$Pr(\Delta P_t^{AA} \neq 0)$	$Pr(\Delta P_t^{DL} \neq 0)$	$Pr(\Delta P_t^{SW} \neq 0)$	$Pr(\Delta P_t^{UA} \neq 0)$
$(LOADDEV_t^{AA})^2$	−0.312	2.291*	−3.063**	1.409*
$(LOADDEV_t^{DL})^2$	0.0267	−0.528	−0.543	−0.617
$(LOADDEV_t^{UA})^2$	−0.078	−0.43	0.49	−0.327
$\Delta P_{t-1}^{AA} \neq 0$	0.405***	0.051*	0.087***	0.146***
$\Delta P_{t-1}^{DL} \neq 0$	0.029	0.461***	0.048**	0.039
$\Delta P_{t-1}^{SW} \neq 0$	0.012	0.049	0.383***	0.069***
$\Delta P_{t-1}^{UA} \neq 0$	0.064**	0.001	−0.017	0.279***
R^2	0.33	0.34	0.26	0.34

Table 3 summarizes the results from regressions of an airline's price deviation on lags of load and price deviations of all airlines. In addition to showing a high degree of fit, our results suggest that competition matters for dynamic pricing: some airlines do respond systematically to the situation of their competitors. For example, all else equal a 10% increase in AA's fares above their usual level leads on average to a statistically significant 0.7% price hike by DL. In terms of capacity, UA responds to a 10-percentage-point increase in the load factor of AA flights by raising its own fares by 2.3%. In other words, if AA flights are already close to full, it represents an opportunity for UA to take advantage of the weakened competition (lower unfilled capacity) by raising prices. More generally, we clearly notice considerable heterogeneity in pricing strategies.

To confirm the results, we focus on price changes instead of levels, in line with Boivin et al. (2012). As is evident from Table 1 and the high coefficient on lagged fares, prices are not changed on a continuous basis by airlines. To disentangle price stickiness from other effects, we define a dummy variable equal to one if a flight's price has been changed by airline *i* since the day before, zero otherwise ($Pr(\Delta P_t^i \neq 0)$). This dummy is then regressed on lagged values of price changes by competitors (ΔP_{t-1}^{-1}) as well as the squared load factor deviation ($(LOADDEV_t)^2$): if an airline's load factor is far from its usual level, one may argue that it should make the company more likely to re-price.

Results in Table 4 generally confirm our findings. For example, UA seems very responsive to price movements of competitors: a price change by AA today makes it almost 15% more likely that UA will modify its fare tomorrow. Delta's decision to change its fares, on the other hand, seems little affected by the pricing decisions of competitors. An airline's load factor deviation seems to have no significant impact on the timing of fare changes.

5. Conclusion

We have described the dynamic pricing patterns observed on the New York–Chicago corridor. We collected high-frequency price and load-factor data for the biggest players on this route. Our findings suggest that the evolution of prices depends not only on the number of days remaining until the flight, but also on the number of seats remaining. Moreover, some of the airlines we study react to competitors' prices and load factors. Specifically, we find

[10] We show this only for American Airlines. Results for the other airlines are similar and so unreported.

[11] Since a given airline may have multiple flights leaving between 6:30 am and 8:30 am, we average fares and load factors.

[12] We aggregate our data up to a daily frequency. More specifically, if there are multiple observations for the same flight at a given extraction date, we compute average load factors and prices for that day. This leads us to have a single observation per extraction date, and makes it possible to use lagged values.

that even after controlling for the number of days remaining until the flight and the number of seats remaining, some airlines increase their prices as their rivals' remaining available seats disappear.

References

Boivin, J., Clark, R., Vincent, N., 2012. Virtual Borders. Journal of International Economics 86 (2), 327–335.

Button, K., Vega, H., 2007. The uses of the "temporal-fares-offered curve" in air transportation. Journal of the Transportation Research Forum 46, 83–100.

Escobari, D., 2007. Systematic Peak-load Pricing, Congestion Premia and Demand Diverting: Empirical Evidence. San Francisco State University, Mimeo.

McAfee, R.P., te Velde, V.L., 2006. Dynamic pricing in the airline industry. In: Hendershott, T.J. (Ed.), 2006. Handbook on Economics and Information Systems, vol. 1. Elsevier, New York.

Nakamura, E., Steinsson, J., 2008. Five facts about prices: a reevaluation of menu cost models. Quarterly Journal of Economics 123, 1415–1464.

Puller, S., Sengupta, A., Wiggins, S.N., 2009. Testing Theories of Scarcity Pricing in the Airline Industry. NBER Working Paper 15555.