Does Scarcity Drive Intra-Route Price Dispersion in Airlines?

Steven L. Puller, Anirban Sengupta, and Steven N. Wiggins[1]

July 2015

Abstract

We use ticket transaction data to investigate the economic forces causing passengers on the same route and airline to pay substantially different fares. First, we show that a simple set of ticket restrictions accounts for the large majority of price dispersion on a carrier-route. Next, we take this basic pricing structure as given and measure how fares change with the scarcity of seats available on a particular flight. Using the models of Dana (1999b) and Gale and Holmes (1993), we test straightforward comparative static predictions about the relationship between flight load factors and fares. We find that scarcity only modestly affects fare levels and has virtually no effect on fare dispersion.

JEL codes: L1, L8, L9

[1] *Puller:* Texas A&M University and NBER, email: puller@econmail.tamu.edu. *Sengupta*: email: anirban.sengupta@gmail.com. *Wiggins:* Texas A&M University, email: swiggins@tamu.edu. We thank Volodymyr Bilotkach, Steve Berry, Severin Borenstein, Diego Escobari, Li Gan, Jim Griffin, Mike Mazzeo, Claudio Piga, Nancy Rose, Jim Dana, Adam Shapiro, and Manuelita Ureta for helpful discussions, and many seminar participants for useful comments. Manuel Hernandez provided excellent research assistance. An earlier version of this paper was circulated under the title "Testing Theories of Scarcity Pricing in the Airline Industry".

1. Introduction

It is well-known that airline fares exhibit substantial price dispersion. Previous research has studied variation in dispersion *across* routes and investigated how it is influenced by changes in competition.[2] This focus on dispersion across routes is driven in part by the fact that the standard data source – Databank 1A/1B – only contains quarterly data at the airline, route level, which precludes a detailed investigation of intra-route price dispersion.

This paper uses a new source of data to investigate the causes of *within route* price dispersion — that is, the factors that lead to differences in the fares paid by individual passengers on a route. We study the influence of ticket characteristics and the scarcity of seats on differences in mean fares and fare dispersion *within a carrier-route*. There is good reason to focus on a different type of price dispersion from much of the existing literature -- within carrier-route fare dispersion is empirically larger than variation in cross-route dispersion. For example, it has been shown that the amount of dispersion *within* a route is roughly twice as large as the dispersion *across* routes.[3] The amount of price dispersion on a route is substantial; as we show below, a typical Gini coefficient is 0.274, indicating that two tickets chosen randomly among passengers on the same route vary on average by 55% of the mean fare.

The analysis below investigates two underlying factors affecting fare differences. The first factor is ticket characteristics, or what we call the underlying institutional structure of airline pricing. Our data contain ticket characteristics such as refundability, advance purchase restrictions, and travel and stay restrictions. We show how these underlying ticket characteristics substantially influence ticket prices and are a key source of fare differences and fare dispersion. This basic structure is often thought to be associated with "price discrimination", however we do not attempt to formally test such theories directly. Rather, our primary goal is to take this basic pricing structure as given and determine how this structure is influenced by changes in the scarcity of seats on a flight.

[2] This research includes, for example, Gaggero and Piga (2011), Gerardi and Shapiro (2009), Borenstein and Rose (1994), Stavins (2001), and others.
[3] Borenstein and Rose (2014) compare the within route fare dispersion to cross-route dispersion after controlling for route distance for the years 1979-2007.

This second factor – the scarcity value of a seat – has been argued to be a potentially large driver of fares. Theories put forth by Dana (1999b) and Gale and Holmes (1993) postulate that airlines charge different customers different fares because of scarcity driven differences in either the opportunity cost of providing a seat or the marginal value of a seat. Dana argues that scarcity pricing is linked to differences in *realized* load factors across flights, and can lead to substantial fare differences. Gale and Holmes argue that scarcity pricing is linked to differences in *expected* load factors across flights, also leading to fare differences.

The models by Dana and Gale/Holmes provide a set of comparative static predictions that we test using a unique dataset of airline ticket transactions. As we describe in detail below, these models predict that flights differing in expected and realized load factors will have different allocations of certain types of tickets, different average fares for the tickets sold, and different levels of price dispersion.

In Dana's model, airlines know the possible realizations of demand for a particular flight and offer seats at different prices for the same flight. The lowest priced tickets will sell for all flights while higher priced seats will sell only when realized demand is high, with the higher prices reflecting the cost of providing seating capacity that will not always be sold. Dana predicts that as *realized* load factors rise, conditional on expected load factor, higher priced seats will constitute a larger percentage of seats actually sold, the average fare of seats sold will rise, and there will be more price dispersion. Dana's theory also predicts that if the bookings on a flight are unusually high at a given time before departure, say 7 or 14 days, then the share of high-priced seats will be higher over the remaining booking period as will the average price of tickets sold.

The Gale/Holmes model makes predictions about the types of flights on which discount, advance purchase seats will be sold. In their model, airlines offer advance purchase discounts in order to shift customers who are more willing to travel in an off-peak period to the flights that fly off-peak. The model predicts that airlines will offer more discount advance purchase seats on off-peak than on peak flights.

Our paper has a simple goal – to test comparative static predictions from these two theoretical models that are often invoked to model intra-route price dispersion in airlines. Our data permit us to examine the relationship between fares and ticket characteristics, and then how scarcity-based factors contribute to differences in mean fares and fare dispersion. Our data consist of transactions through one of the major computer reservation systems matched with data on ticket characteristics from a second major computer reservation system. For each itinerary, our data include information on the carrier-route, fare, ticket restrictions, flight numbers, and dates of purchase, departure and return. The flight level data permit us to calculate various measures of load factors, including expected load factors, realized load factors at departure, and realized load factors at intermediate points in the booking process.

There are several advantages of our data that enable us to investigate the issues described above. One key advantage is that our data contain ticket characteristics, which permit us to analyze the influence of these characteristics on pricing. These data also permit us to examine the quantity allocations of tickets with different types of restrictions on different flights. A second advantage is that our load factor data permit us to investigate the linkage between scarcity of seats on given flights and the quantity allocation of different types of tickets on those flights. The final advantage is that we can investigate how changes in both expected and realized load factors affect the level and dispersion of ticket prices. Because we know dates of both purchase and departure, we can investigate how variation in realized load factors during the booking process, say 7 or 14 days prior to departure, affects both prices and quantities associated with subsequent transactions.

Our key contribution is to investigate the extent to which fare differences on the same carrier-route are driven by scarcity, as predicted by the Dana and Gales/Holmes models. Our empirical strategy is to identify a set of flights that have very low ex ante probability of reaching capacity and ex post have a relatively small load factor at departure. The seats on these flights have very low scarcity value because the probability that the last seat will be filled is small. These baseline flights should have fares that reflect the basic institutional structure of fares (e.g. refundability and travel restrictions), but the fares should *not* reflect scarcity. We compare the tickets sold on the baseline flights to those sold on flights with higher expected and realized load factors, using the two theories of scarcity pricing to provide predictions. The models of Dana and

Gale/Holmes provide seven comparative static predictions about the relationship between the types of tickets sold on specific flights and the load factors of those flights.

We find that scarcity has a modest impact on fares in some dimensions and very little impact in other dimensions. First, we find that flights with higher levels of expected or realized load factor have only a slight tendency to sell fewer restricted tickets. Second, the average fares on flights that are unexpectedly full are somewhat higher, especially when focusing on tickets sold in the days just before departure on flights that had been filling up unexpectedly quickly. Third, the dispersion of fares is surprisingly unrelated to scarcity; fare dispersion is virtually identical on the set of baseline flights as on flights with higher expected and/or realized demand. These findings suggest that scarcity has virtually no ability to explain the large amount of dispersion in fares paid by passengers travelling on the same route and the same airline.

The remainder of the paper is organized as follows. Section 2 reviews theories of airline pricing and describes how this paper contrasts with the existing empirical literature on airline price dispersion. Section 3 describes our data. Section 4 presents the empirical tests and Section 5 concludes.

2. Predictions of the Theories of Scarcity Pricing in Airlines

Theories of scarcity pricing have been developed to explain differences in the fares sold for tickets on flights with varying demand characteristics. As contributions to the theoretical literature, these theories are developed for stylized descriptions of the airline industry that necessarily abstract from the complex institutional structure of airline pricing. That institutional structure includes attaching restrictions to tickets such as non-refundability, advance purchase requirements, and restrictions on the day of travel or length of stay. It is widely believed that at least a portion of this institutional structure reflects price discrimination, but we do not attempt to formally test such theories directly. Instead we take this pricing structure as given and assess how scarcity influences fares and ticket allocations.

Therefore, our empirical strategy is to evaluate these theories of scarcity pricing in the context of this basic institutional structure of airline tickets. As a preliminary step, we quantify the basic institutional structure and show associations between ticket characteristics and pricing. Then we conduct the primary empirical analysis by testing the impact of rising scarcity on the allocation of tickets, pricing of those tickets, and fare dispersion in the context of this institutional structure.

We draw upon two major groups of theoretical models of scarcity pricing developed for airlines. One group of models predicts how pricing is determined by demand uncertainty and is formalized by Dana (1999a, 1999b). The second group of models describes pricing in peak and off-peak periods, and is developed in Gale and Holmes (1992, 1993). Both groups of models are built on the highly perishable nature of airline seat inventories, a principle that applies generally to hospitality industries. This perishability means that an empty seat has significant economic value that fully dissipates when the flight takes off.[4]

Dana's analyses build on models originally developed by Prescott (1975) and Eden (1990), which can be used to explain inter- and intra-firm price dispersion when inventories are perishable. In these models, there are capacity costs λ per seat and there may also be a marginal cost, which we will ignore for purposes of exposition. It is easiest to follow Dana's analysis by considering a simple setting where there are two possible demand states -- low demand occurring with probability θ and high demand occurring with probability $(1-\theta)$. Consumers arrive and purchase the lowest priced ticket that is available when they arrive. Firms must set the prices of all tickets before the demand shock is realized.

Dana shows that there exists a pure strategy equilibrium in which firms offer some portion of the available seats priced at capacity cost, $p=\lambda$. These seats sell in both the high and low demand states. In addition, the firms offer another set of more expensive seats priced at $p=\lambda/(1-\theta)$, where $1-\theta$ is the probability of the high demand state which is the only state in which those seats sell. This pricing arises from a zero profit competitive equilibrium in which the expected revenue of each seat equals the marginal costs of capacity. Intuitively, the scarcity cost-based premium "compensates" the firm for the incremental capacity cost of offering some seats of perishable value that sell only some of the time.

[4] Other models have focused on modeling the dynamic pricing of airline tickets as departure date nears and demand shocks are realized, for example Gallego and Van Ryzin (1994).

In this model, firms sell only "low" priced tickets when *realized* demand is low, and sell both "low" and "high" priced tickets when *realized* demand is high. The result is intrafirm price dispersion. This result is achieved without using restrictions or "fencing" devices such as advance purchase discounts or required Saturday night stays. The model in Dana is more general than the two demand state case described here; it generalizes to multiple possible demand states as well as to other forms of market structure such as monopoly and oligopoly. Dana also shows that under forms of market structure other than monopoly, firms compete in price distributions and thus there is intra-firm price dispersion. However, the key testable empirical implications can be seen in this simple form of the model.

In testing this model, it is important to note that Dana's model provides implications for *transacted* tickets and *realized load factors.* For a set of flights with the same ex ante distribution of demand, the set of *offered* fares will be the same. However, the subset of flights with higher *realized* load factor will have different *transacted* fares. Hence testing this theory requires an evaluation of the relationship between transacted fares and realized load factor after controlling for expected load factor.

The Dana model predicts four comparative static relationships between flights with different realized load factor *but the same ex ante distribution of demand*. Each of the following predictions arise from the model. *Prediction 1*: The mean fare of transacted tickets will be higher on flights with high realized load factors; *Prediction 2*: There will be more fare dispersion on flights with high realized load factors; *Prediction 3*: The share of high-priced tickets will be larger on flights with high realized load factors; and *Prediction 4*: When bookings are unusually high for a given number of days before departure, there will be more high-priced tickets sold in the remaining days before departure as compared to flights where bookings are average or below.[5] Although these predictions are obviously related, we test each of these predictions and investigate their ability to organize airline pricing data and explain fare dispersion.

---

[5] The model also predicts more dispersion in routes that have more competition, which is consistent with results from Borenstein and Rose (1994). However, Borenstein and Rose provide a different model yielding dispersion -- a monopolistically competitive model with certain demand. We do not test predictions regarding market structure because we seek to understand the causes of *within* carrier-route dispersion.

As we describe below, we identify flights with similar ex ante distributions of demand by finding flights on the same carrier and route with similar levels of expected load factor.[6] We test these predictions by comparing flights with similar levels of expected demand but different realizations of demand shocks. In particular, we control for expected load factor and test whether flights with higher realized load factors have the characteristics described in the four predictions.

The second group of models that provide testable predictions are developed in Gale and Holmes (1992, 1993) to predict differences in pricing on peak versus off-peak flights. The models use a mechanism design approach to model how a monopolist offers advance purchase discounts. Such discounts are used to divert customers to off-peak flights. In their model, suppose that a monopoly airline has one peak and one off-peak flight. Each consumer prefers either the "peak" or "offpeak" flight, but only learns her preferences near departure; for example, a customer may know she has a business meeting on a certain date but not learn the time of the meeting until shortly before that date. Also assume that customers vary in their disutility of taking the flight that occurs at the "wrong" time (i.e. they have different opportunity cost of waiting time). Gale and Holmes show that in equilibrium the monopolist offers discounted advance purchase tickets on the off-peak flight during the period before uncertainty regarding preferred flight times is resolved. Discounts are not offered on off-peak flights for purchases that are not made in advance, nor are any discounts ever offered on the peak flight. In equilibrium, customers with a low opportunity cost of waiting will purchase the discount advance purchase tickets and this mechanism serves to shift those customers to the off-peak flight. Thus, the off-peak flight consists of two types of travelers: those with a low value of time who purchased at the discount in advance and those with a high value of time who preferred to travel in the off-peak period. The peak flight consists only of passengers who paid regular fare. The main empirical implication is that airlines sell (more) discounted advance purchase tickets on the off-peak flight than on the peak

[6] In principle, one could match on other moments of the demand distribution, but we match on mean to make the strategy tractable.

flight.[7] We classify this model as a "scarcity pricing" model because the lower fares in the off-peak period reflect a lower opportunity cost of service.

The Gale and Holmes model provides three additional empirical predictions to test with our data. The predictions are comparative static relationships between *expected load factor* and *transacted fares.*[8] *Prediction 5*: there will be fewer discount/advance-purchase seats sold on peak flights—that is, flights with a high expected load factor. *Prediction 6*: flights with high expected load factors (peak flights) will have higher average fares than off-peak flights with low expected load factors. *Prediction 7*: flights with higher expected load factor (peak flights) will have less price dispersion because discounts are not offered for those flights.[9]

It is important to recognize that testing these seven theoretical predictions must account for the complex institutional structure of airline pricing. The theoretical models are necessarily simplified. To the extent that these theoretical models operate in the U.S. airline industry, they do so in conjunction with the existing fare structure of airlines. Thus, we do not interpret these models too literally and expect the predictions to fully organize the data on airline ticket transactions. Rather, the analysis below considers whether these theories contribute to a better understanding of overall airline price levels, price dispersion, and the allocation of ticket types across flights. To foreshadow our empirical strategy, we test each of the seven predictions by comparing flights with different levels of expected and/or realized load factor, using a simple regression strategy. Because demand shocks do not play a role in the Gale and Holmes model, the empirical tests of the Gale and Holmes predictions relies on comparing flights with different levels of expected load factor; the empirical model does not include realized load factor as a

[7] In a related paper, Gale and Holmes (1992) allow for uncertainty in the peak period, and find that at least some advanced purchase tickets are sold on the peak flight. Dana (1998) builds upon the advance purchase literature and shows that advance purchase discounts can arise in a perfectly competitive setting.

[8] In their conclusion, Gale and Holmes (1993) suggest that empirical testing should follow precisely the line of research that we do: "The main empirical prediction of this paper is that airlines will limit the availability of discount seats on peak flight. There is anecdotal evidence that airlines do this, but more careful study is needed. Ideally, such a study would employ a flight-level data set that specifies, for a given date and time, the fares paid by all of the passengers on that flight. Unfortunately, this is greater detail than is available in the airline data sets currently in use."

[9] Note that this prediction about dispersion holds in the setting where the timing of the peak is certain ex ante, as in Gale and Holmes (1993). In an extension paper, Gale and Holmes explore the case where the timing of the peak is uncertain (Gale and Holmes, 1992). In that setting, the prediction about dispersion is ambiguous because the peak is not known ex ante by the airline.

covariate. However, the Dana model makes predictions about the characteristics of flights with the same ex ante demand but different realizations of demand. Thus, the empirical tests will compare flights with different realized load factors but similar expected demand. We control for expected demand and test whether coefficients of realized load factor are consistent with the model's predictions. Because the two theories make predictions of comparative static relationships, we use a least squares regression model.

Before turning to our data, it is important to note the differences between our analysis and the existing literature on airline price dispersion. As noted in the introduction, most of the existing literature investigates the causes of price dispersion *across* routes rather than *within* routes. The previous literature derives and tests predictions for the impact of market structure on price dispersion; this literature has implications for the effects of mergers and the nature of competition.[10] In contrast to the previous literature, our paper improves our understanding about the extent to which scarcity is a driver of price dispersion on a given carrier-route, which as discussed above, is larger than the dispersion across routes.

Our paper is complementary to another branch of the literature that uses posted fares to study the evolution of fares over time. McAfee and Velde (2007), for example, draw upon the yield management literature and devise results for dynamic price discrimination and the efficient allocation of seats when airlines are faced with demand uncertainty. They use data gathered from online websites to study the price paths for specific flights as departure nears. They find only weak evidence of dynamic price discrimination. Escobari and Gan (2007) gather and use data from posted minimum prices to test Dana's theories. Alderighi, Nicolini, and Piga (forthcoming) study the yield

[10] Borenstein and Rose (1994) analyze the relationship between price dispersion and market structure using a cross-section of markets. They show an increase in dispersion as markets become more competitive. Gerardi and Shapiro (2009) investigate the effect of competition on dispersion using a panel data set, and in contrast to Borenstein and Rose, find that increases in the competitiveness of a route reduce dispersion. They interpret their findings to suggest that more competition reduces the ability of incumbent carriers to implement price discrimination. Stavins (2001) uses a novel data set on posted prices and a subset of ticket characteristics, namely Saturday night stay-over and refundability, to find evidence consistent with both Saturday night stay and refundability being used as price discriminating devices. Using these data, Stavins corroborates the finding of Borenstein (1989) that an increase in a carrier's share is associated with higher prices, and the finding of Borenstein and Rose (1994) that increased competition on a route is associated with higher price dispersion.

management process of a European low cost carrier using detailed webscraped data from the carrier's website. Dobson and Piga (2013) study the effects of mergers on the fare structures chosen by airlines.[11]

This paper contributes to a growing literature that seeks to understand the determinants of intra-route airline pricing. Important contributions to this literature exploit structural models of customer demand and airline pricing to investigate price dispersion. Williams (2013) studies the interaction between intertemporal price discrimination and stochastic demand pricing. He develops a model with consumers making a static choice and a monopoly airline facing a dynamic pricing problem in which aggregate demand is uncertain and the types of customers who purchase vary as departure date nears. Using daily data on lowest available fares and seat availability, he finds that dynamically adjusting fares as departure approaches in response to observed demand shocks increases airline revenues and leads to a more efficient utilization of capacity. Lazarev (2013) studies the effect on welfare of intertemporal price discrimination as compared to alternative settings such as allowing airline ticket resale via secondary markets. Using a model of a monopoly airline setting fares to forward-looking consumers, he finds that free ticket resale that eliminates price discrimination would decrease the consumer surplus of leisure travelers and increase consumer surplus of business travelers. Escobari (2012, 2009) studies the dynamics of pricing as inventory falls and departure nears, and he investigates the frequency of discount seats on peak flights. These papers provide insights on the strategic nature of pricing. In particular, subject to modeling assumptions, these papers sharpen our understanding of the interaction between price discrimination and demand uncertainty in settings where the supply or demand-sides are dynamic.

[11] Related empirical work has studied other pricing and load factor phenomena. Sengupta and Wiggins (2014) study the effect of on-line sales on pricing. Dana and Orlov (2014) investigate whether the increased use of internet booking leads airlines to increase capacity utilization. Goolsbee and Syverson (2008) investigate the effect of the threat of Southwest entry on incumbent carrier pricing. Forbes (2008) estimates the effect of delays on fares. Other research has studied the effect of airline bankruptcy or financial distress on pricing, including Borenstein and Rose (1995), Busse (2002), and Ciliberto and Schenone (2012). Berry and Jia (2010) explore a variety of demand and supply side explanations for reduced airline profitability in the last decade despite increases in both load factor and passenger miles flown. Lederman (2008) studies the impact of frequent-flyer program on hub premia.

3. Data

3.A. Tickets in Our Sample

We use the population of all transactions conducted via one of the major computer reservations systems (CRSs) for the fourth quarter of 2004. This CRS handles transactions for all major channels of ticket purchases: tickets purchased through travel agents, several major online travel sites, and directly from airlines, including their web-based sales. These data comprise roughly one-third of all domestic U.S. ticket itineraries. For each itinerary, we observe the fare and date of purchase. In addition, for each flight segment of a given itinerary, we observe the travel date, origin and destination, airline, flight number, and cabin class of service.

Following Borenstein (1989) and Borenstein and Rose (1994), we analyze the pricing of coach itineraries with at most one stop-over in either direction. Specifically, starting with our set of all transactions, we exclude itineraries involving travel in the first class cabin. In addition, we exclude itineraries with open-jaws and circular trip tickets. We analyze the prices of round-trip itineraries; we double the fares for one-way tickets to obtain comparability. This study includes tickets for travel on American, Delta, United, Northwest, Continental and USAir. These constituted all airlines that in 2004 carried at least 5% of U.S. domestic customers with the exception of Southwest for whom we have only limited data. We analyze tickets for travel in the fourth quarter of 2004 excluding travel on Thanksgiving weekend, Christmas, and New Years.

We restrict our analysis to large routes. To choose these routes, for each of the six carriers, we stratify the sample to include the largest routes for each carrier with varied market structures.[12] The routes are listed in Table 1. We include tickets by any of the six carriers listed above that serve any of the routes listed. One consequence of choosing large routes is that the sample consists largely of routes from airlines' hubs—though this should not pose an issue for testing the general theories of airline pricing under investigation.

[12] In particular, we choose the largest routes within each category of market structure – monopoly, duopoly, and competitive – using the market structure definitions from Borenstein and Rose.

3.B. Ticket Characteristics

Because we wish to observe ticket characteristics that impact a traveler's utility, we merge our transaction data to a separate dataset that contains ticket-level restrictions. Travel agents' computer systems can access historical data on posted prices for up to a year. We collected additional data on restrictions from a local travel agent's CRS. The historical archive contains a list of fares/restrictions for travel on a specified carrier-route-departure date. For each archived fare, we collected information on carrier, origin and destination, departure date from origin, fare, booking class (e.g. first class or coach), advance purchase requirement, refundability, travel restrictions (e.g. travel can only occur on Tuesday through Thursday), and minimum and maximum stay restrictions. We merged these data to the transaction data.

The matching procedure is described in detail in the Appendix. Briefly, we match our transacted itineraries to the archive of fares/restrictions based upon carrier, departure date, fare, consistency between purchase date and a possible advance purchase restriction, and tickets where travel dates were consistent with the posited travel and stay restrictions. We keep matches if the tickets meet these criteria and the fares are within two percent of each other. If a transaction ticket matches multiple posted fares, we take the closest match based on fare. Further details are in the Appendix.

We are able to match 36 percent of the observed transactions to data from the travel agent's CRS. However, there do not appear to be substantial differences between the matched and unmatched transactions. Table 2 compares means of all transactions to those that we successfully match to fare characteristics. The data in the table indicate only modest differences between the matched and unmatched transactions. The unmatched transactions tend to be slightly lower priced tickets – across all the carrier-routes, the matched tickets average $424 while all tickets average $415. In general, ticket characteristics are very similar between matched and all transactions. We analyze whether these unmatched tickets tend to come from particular segments of the price distribution. In Figure 1, we plot kernel density estimates of prices. Although we tend to match fares that on average are slightly higher, we are able to match fares from various parts of the fare distribution.

3.C. Measuring Realized and Expected Load Factors

The theories of scarcity pricing discussed above make predictions based upon both realized and expected load factors. Because our transaction data include the date of purchase, we are able to calculate realized load factor either at departure or at any point in time before departure.

In order to calculate realized load factor, we combine three sources of information. The Official Airline Guide contains information on the number of available seats on each flight in our sample. Our transaction data include flight-specific purchases for roughly one-third of all ticket transactions.[13] We use additional Bureau of Transportation Statistics (BTS) data to ensure an unbiased estimate of the total number of itineraries on a particular flight, including those sold through other computer reservation systems. The BTS reports monthly data on tickets sold by carrier-route, which permits us to calculate the exact share of tickets that we observe for a carrier-route in our CRS data. We then scale up the observed coupons on a particular flight by the inverse of that observed share to obtain an unbiased estimate of realized itineraries sold for a particular flight.[14] Realized load factor is calculated as the number of total itineraries sold divided by the number of available seats. As we note above, we can calculate the realized load factor at departure or at any specified number of days before departure. Although load factor will be measured with error, our methodology implies the measurement error will have zero mean at the route-airline level. Load factor should also be unbiased at the flight level because the CRS share is unlikely to vary systematically for particular flights or days of the week.

Calculation of expected load factor is more difficult, particularly because it involves expectations and because we are limited to twelve weeks of data. One could imagine different ways to measure an airline's expectation of the load factor for a future flight. Our definition of expected load factor accounts for the ability of airlines to predict that certain timeslots and certain weeks of the year have systematically different demand.

[13] We account for all customer itineraries sold for a given flight, whether that itinerary is for the route in question or if the itinerary is a "connecting passenger" on that flight traveling on a different route.

[14] For example, for American Flight 301 from New York La Guardia (LGA) to Chicago-O'Hare (ORD) on October 11, 2004, we measure the number of seats (129) and the number of tickets sold through the CRS that include this flight on the itinerary (26). Because American sells 36 percent of its tickets for direct service between LGA and ORD through our CRS, we calculate the realized load factor to be 55 percent (=(26/0.36)/129).

Our metric of expected load factor is motivated by two insights about air travel. First, the demand for air travel has very systematic patterns over the course of a day, the week, and the year. In particular, demand varies by time of day – this is driven by the networked nature of air travel, particularly among the hub-and-spoke carriers in our sample. Also, demand is systematically different on weekdays than on weekends. Finally, there are seasonal patterns so that demand varies from one week to the next. The second insight is that these patterns can vary by both the carrier and the route served. Therefore, we create a metric of expected load factor that accounts for daily, weekly, and seasonal patterns and allows for these patterns to vary by both the carrier and route.

We identify peak and off-peak times for airline demand based upon particular departure timeslots during the week and weeks of the year where seat demand varies. Specifically, we define seven timeslots during the week that have systematically different average load factors in our sample: weekdays 1-5am, weekdays 6-9am, weekdays 10am-1pm, weekdays 2pm-7pm, weekdays 8pm-midnight, Saturdays, and Sundays. We also allow for differences in expected demand across weeks. Any individual flight is placed in its appropriate category of carrier, route, timeslot, and week-of-year. Then we define the flight's expected load factor as the sample average actual load factor across all flights for a given carrier-route-timeslot-week-of-year. For example, we use the sample average of the actual load factor of all flights on American on the LaGuardia-O'Hare route that departed on weekdays between 10am and 1pm in the second week of October as the metric of expected load factor for the flights matching those criteria.

The theories of scarcity pricing provide comparative static predictions about both fare levels and dispersion on flights with high versus low expected and realized load factors. Dana focuses on realized load factors *conditional on the expected load factor*, and indicates that any flight with higher realized load factors will sell more high-price tickets. Testing this theory requires one to sort flights by expected load factor, and then calculate differences in realized load factors within these groupings of expected load factor. Gale and Holmes focus on expected load factor and posit that off-peak flights will have more advance purchase restrictions and lower average fares. The Gale and Holmes theory differentiates across levels of expected load factor while remaining silent on realized load factor.

Accordingly, we divide individual flights at the carrier-route level into expected load factor quartiles. First, we separate flights into quartiles using the carrier-route-timeslot-week-of-year metric of expected load factor. We denote these groups as expected to be "Full", "Medium-Full", "Medium-Empty" and "Empty" as illustrated in the columns of Figure 2. Within these quartiles that contain flights of similar *expected* load factors, we then calculate quartiles of *realized* load factors for individual flight departures within a carrier-route. This procedure differentiates among realized load factors based on within carrier-route differences *within a category of flights with similar expected load factors*, which is required to test Dana's theory. In some of the analyses below, we also measure realized load factor 7 days prior to departure to examine how unexpectedly high (or low) load factors during the booking process affects prices and quantities for subsequent transactions. Figure 2 shows the resulting matrix of expected and realized load factors. Of particular importance in our analysis are flights where there is arguably little scarcity at work, i.e. flights that are expected-to-be-empty that are realized-to-be empty (hereafter "Empty/Empty" flights).

An example illustrates. All of American's flights from La Guardia to O'Hare are grouped into 4 categories of expected load factor using the carrier-route-timeslot-week-of-year metric. Consider the flight American 301 on October 11, which departs on weekdays at 6:00am. We group this flight with all other American flights on the La Guardia to O'Hare route departing on a weekday between 6 and 9am during the week of October 11, and calculate the average realized load factor. It turns out that this average load factor is relatively low compared to other timeslot-week-of-year average realized load factors on the LGA-ORD route for American – the average load factor is in the second quartile. Therefore, all American flights on the route LGA-ORD departing on a weekday 6-9am during the week of October 11 are classified as Expected to be "Medium-Empty". We then categorize the individual departures of expected to be "Medium-Empty" flights on American's route LGA-ORD into four quartiles of Realized load factor at departure. American's Flight 301 on October 11 with a realized load factor of fifty-five percent is among the lowest load factor flights of those in the "Medium-Empty" expected load factor; therefore tickets on this flight are categorized as "Expected to be Medium-Empty and Realized to be Empty." This calculation is based upon realized load factor *at departure*. We use a similar approach to calculate an analogous categorization based upon realized load factor as of seven days before departure. To calculate our measure of load factor as of seven days before departure, we categorize Flight 301 on October 11 based on its realized load factor at that time (October 4), and then considered transactions between October 5 and departure.

These categories permit us to calculate how expected and realized load factors influence ticket prices and quantity allocations while controlling for carrier-route differences. It also permits us to investigate fares and quantities in settings where there is little scarcity -- Empty/Empty flights -- and then to compare fares and quantities for flights with higher load factors, fully controlling for carrier-route differences. The analysis below examines three dimensions of pricing and quantity allocations: (1) the allocation of tickets to pricing groups (bins) according to refundability and restrictions, (2) overall average prices of tickets sold, and (3) Gini coefficients to measure dispersion.

3.D. Summary Statistics

Our sample includes all combinations of carriers and routes listed in Table 1 for which we have at least 1000 itineraries, leading us to study 119 different carrier-routes. Summary statistics of the transaction data that we include in our sample are shown in the first column of Table 2. In some empirical tests below, we use all of the transactions while in others we use only transactions that can be matched to ticket characteristics. Table 2 reports summary statistics for both the entire sample and those itineraries with ticket characteristics.

Among the entire sample of 612,903 itineraries, fares average $415 for roundtrip travel. Most tickets are purchased in the days shortly before departure; the fraction of tickets purchased within three, six and thirteen days before departure are 28%, 42% and 62%, respectively. The day of the week with the most initial departures is Monday and the day with the fewest departures is Saturday. We measure ticket characteristics for 221,895 (or 36%) of these itineraries. Generally the matched tickets are very similar to the sample as a whole. The additional information on ticket characteristics indicates that 26% of tickets are refundable, 38% include some form of travel restriction, 20% include a minimum stay restriction (e.g. 1 night stay), and 15% include a maximum stay restriction (e.g. 30 days).

Panel B of Table 2 reports summary statistics on price dispersion for flights with different levels of expected load factor. The Gini coefficient is remarkably similar across all combinations of quartile of expected and realized load factor. This is suggestive evidence that scarcity does not drive dispersion; we test this formally in section 4.D.

4. Empirical Results

4.A. Airline Pricing Institutions: Ticket Prices and Restrictions

Our first set of regressions illustrates the basic institutional structure of airline fares by showing the close linkage between fares and ticket restrictions. The data contain four different types of restrictions: refundability, restrictions on the dates of travel, restrictions on the length of stay, and whether the ticket was for roundtrip travel. Table 3 presents two different groups of regressions where the natural log of fares are regressed on these restrictions. In these regressions, an observation is an individual travel itinerary. As we note above, our goal is not to explain the determinants of this basic structure, but rather to describe this structure and then to estimate whether it changes with the scarcity of seats available on a particular flight.

The first column of Table 3 presents results when the log of fares is regressed on ticket characteristics: refundability, a roundtrip indicator, and indicators for whether the ticket included travel and stay restrictions.[15] The regression contains carrier-route fixed effects, so the coefficients describe how ticket characteristics explain within carrier-route variation in fares. In addition, we include fixed effects for the week of year of travel to allow for changes in fuel prices that could shift the distribution of fares. The results show that this basic institutional structure explains a large share of the variation in fares paid. The observed characteristics explain two-thirds of the variation in log fares (the R^2 is

[15] We also could use advance purchase restrictions or days in advance that a ticket was purchased in order to explain the variation in fares. In unreported regressions, those factors also are statistically significant and explain additional variation. We do not use advance purchase in this specification because our later empirical tests separately distinguish whether a ticket was purchased in advance.

0.66).[16] The results also show substantial explanatory power for each of the observed ticket characteristics. Passengers who purchase refundable tickets pay a 56% premium. Tickets with restrictions on the days of travel or the length of stay are sold at prices 33% and 14% lower, respectively. Finally, a roundtrip ticket is sold at a 20% lower fare than two one-way tickets for the same travel.

In the second column of Table 3, we show that this strong relationship between characteristics and fares exists *even on flights with limited scarcity*. We estimate the same model as above except that we restrict the sample to itineraries that involve travel on "Empty/Empty" flights, as defined in section 3.C. These are flights where arguably scarcity of seats plays a limited role. The results in column 2 show that the ticket characteristics have a very similar relationship with fares on Empty/Empty flights as they do on all flights in the sample.

The third column in Table 3 shows that a simple grouping of fares into three categories captures the major dimensions of fare variation. We define three different "Groups" of tickets which are defined based on ticket characteristics. "Group 1" constitutes fully refundable, unrestricted tickets; "Group 2" represents nonrefundable tickets that do not have travel or stay restrictions; and "Group 3" represents nonrefundable tickets that include travel and/or stay restrictions. Group 1 mean fares are $632, the Group 2 mean is $440, and the Group 3 mean is $281. The relative shares of Groups 1, 2, and 3 tickets are 26%, 32%, and 42%, respectively.

A key advantage of this grouping is that it corresponds well to the theories -- Group 1 consists of high priced tickets, Group 2 medium, and Group 3 low priced tickets. The higher fares in Group 1 also tend to correspond to tickets that are *not* sold "in Advance". Only 9% of Group 1 tickets have explicit advance purchase restrictions and only 4% are actually purchased more than 21 days in advance. In contrast, 75% of Group 3 tickets have advance purchase restrictions and a disproportionate number are sold several weeks before departure. For Group 1, Gale and Holmes predict that the portion of these seats sold on peak (i.e. high expected load factor) flights should be higher than

[16] This explanatory power is not driven primarily by the fixed effects; the fixed effects alone explain only 35% of the variation in fares.

on low expected load factor flights. Dana's model predicts that, holding expected load factor constant, the Group 1 share should increase with increases in realized load factor.

A second advantage of these groupings is that they capture a large share of the variation in prices. The third column of Table 3 shows a regression of fares on Groups 1-3 and the same fixed effects included in the first column. Relative to the excluded category of Group 2, tickets in Group 1 are sold at a 51% higher price and Group 3 tickets are sold at a 45% lower price. These groups (in addition to the carrier-route fixed effects) explain 63% of the variation in log fares. The results show that this grouping captures a large share of the fare variation associated with the institutional structure of airline fares. We now test how scarcity works within this institutional structure to influence the fares paid by customers.

4.B. Methodology for Testing Predictions about Scarcity Pricing

We test how scarcity pricing operates within the institutional structure of airline pricing in several different ways. The theory discussed above provides comparative static predictions about several metrics of transacted tickets, including the share of different types of tickets, average prices of those tickets, and price dispersion. Each of these metrics is predicted to vary with respect to expected and/or realized load factor.

In order to facilitate testing of the seven predictions from the models, we utilize a common empirical framework. Each outcome – ticket allocations, average prices, and dispersion – is modeled as varying in expected and realized load factor from a baseline level. Our baseline is the set of flights that are both expected to be in the lowest quartile and realized to be in the lowest quartile of load factor, which we refer to as Expected to be Empty and Realized to be Empty. We view the flights that are "Empty/Empty" as ones that reflect the basic institutional structure of pricing but should not reflect scarcity-based pricing because those flights are not expected to be in high demand and those flights did not experience unexpected positive shocks to demand.

We test the seven predictions in a regression framework. An observation is a cell of the 4x4 matrix of expected and realized load factors in Figure 2, which we create separately for each carrier and route combination. Our basic empirical framework is to regress each measure of outcome (allocation, prices, and dispersion) on categorical variables for each quartile of expected and realized load factor:

$$\begin{aligned} Y_{ij} = \delta_0 + \delta_1 ExpectedOrRealizedEmpty\_ButNotBoth_{ij} \\ \beta_1 ExpectedMediumEmpty_{ij} + \beta_2 ExpectedMediumFull_{ij} + \beta_3 ExpectedFull_{ij} + \\ \gamma_1 RealizedMediumEmpty_{ij} + \gamma_2 RealizedMediumFull_{ij} + \gamma_3 RealizedFull_{ij} + \varepsilon_{ij} \end{aligned} \quad \textbf{(1)}$$

Y_{ij} is the outcome variable (i.e. allocation of a ticket type, average fare, and dispersion) for cell i of carrier-route j. Each carrier-route j has 16 cells corresponding to all combinations of the 4 quartiles of Expected and Realized load factor. We specify the model so that the coefficient δ_0 captures the baseline outcome on flights that are expected-to-be and realized-to-be "Empty". As we note above, we view this baseline as the set of flights in which scarcity is not present with respect to either expected demand (Gale and Holmes) or realized demand (Dana). To do so, we create another variable that is "Expected to be Empty or Realized to be Empty *but not both*". This variable captures effects where there is arguably only a modest effect of scarcity because load factors are in the lowest quartile in one dimension but not the other. Given this specification, δ_0 captures the bottom-right cell in Figure 2.[17] The coefficients β_1,β_2,β_3 capture the effects of higher expected load factor, i.e. moving columns to the "left" in Figure 2. The coefficients $\gamma_1,\gamma_2,\gamma_3$ capture changes in the outcome as realized load factor rise (conditional on expected load factor), i.e. moving "up" to higher rows in Figure 2. And, finally δ_1 has no theoretical prediction – it merely is used to separate the effects of being either expected or realized to be empty from our baseline of expected *and* realized to be empty.

All of our empirical tests can be carried out by examining particular specifications within this general model. Gale and Holmes make predictions about the effect of expected load factor without addressing the role of realized load factor, so we estimate models only with the three quartiles of expected load factor (and the constant). Dana

[17] In Figure 2, δ_0 captures the bottom-right cell, and δ_1 captures the right column and the bottom row, excluding the bottom-right cell.

makes predictions about realized load factor conditional on expected load factor, so we include all covariates in the general model.

All specifications are estimated with least squares and we report standard errors that are clustered at the carrier-route level. In some specifications, we include carrier-route fixed effects and in others we do not, for reasons that we describe below.

4.C. Scarcity and the Allocation of Seats Across Groups

We begin by examining how scarcity—expected and realized load factor—influences the shares of different types of seats as predicted by Gale and Holmes (Prediction 5) and Dana (Predictions 3 and 4). Later in sections 4.D and 4.E, we examine other channels through which scarcity could influence average prices paid (Predictions 1 and 6) and price dispersion (Predictions 2 and 7).

To examine the relationship between scarcity and shares, we begin by separating individual flights into expected load factors at the carrier-route level. These categories are used to test the Gale/Holmes predictions about peak and off-peak flights. Because Dana's model yields predictions about flights with higher realized load factor after conditioning on expected load factor, we then sort within each of these expected load factor groupings to create quartiles of realized load factors. These procedures result in a grouping of flights that populates the cells illustrated in Figure 2. We then calculate shares of tickets for Group 1, Group 2, and Group 3 in each of these cells at the carrier-route level.

Table 4 examines the relationship between load factors and shares of particular types of tickets during particular periods of time before departure. Gale and Homes predicts there are fewer tickets with "Advance Purchase" restrictions tickets sold on "peak" flights (Prediction 5). This implies that for high levels of expected load factor, one should observe fewer Group 3 tickets sold in advance. (Recall that Group 3 tickets are highly restricted tickets). Column 1 of Table 4 begins by considering the share of Group 3 tickets sold more than 21 days in advance as a fraction of total ticket sales. The regressors in column 1 include quartiles of expected load factor. The results do not support the Gale and Holmes prediction -- there is not a statistically significant relationship between the various categories of expected load factor and the share of

Group 3 tickets sold more than 21 days in advance. The coefficients for the second, third, and fourth quartiles of expected load factor are negative and are close to -.01, which indicates that the point estimates represent approximately a one percentage point lower allocation of tickets to Group 3. This statistically insignificant effect compares to the baseline of 16% of tickets that are Group 3 tickets sold 21+ days in advance on flights that are expected to be "empty" (i.e. 1st Quartile). An F-test fails to reject the hypothesis that the coefficients of expected to be quartiles 2, 3, and 4 are simultaneously equal to zero (p-value = 0.32). Column 2 adds carrier-route fixed effects to allow for Expected to be Empty flights to have different Group 3 shares across carriers and routes. Results are quantitatively very similar which is not surprising if factors driving the allocation of advance purchase tickets across flights are largely independent of the route and carrier.

The third and fourth columns of Table 4 tests the sensitivity of this result to the time period used to define an "Advance Purchase". These specifications analyze the share of Group 3 tickets sold more than 14 days prior to departure. We obtain point estimates that are slightly smaller than the point estimates under the 21 day specification above. Again, there is no statistically significant relationship between expected load factor and the share of advance purchase restricted tickets.

The implication of these results is that there is no strong support for the key prediction of Gale and Holmes that tickets with advance purchase restrictions will be allocated disproportionately to "off-peak" flights with low expected load factors.

We now use this same approach to address the quantity predictions provided by Dana's model. Dana's model predicts that cheaper seats sell out on flights when realized demand is high, after controlling for expected load factor. Accordingly, the share of tickets sold that are "high priced" should be higher on flights with a higher realized load factor, conditional on expected load factor. This prediction focuses on how these shares should differ over the period near departure when many of the transactions occur and a plane fills up. Therefore, we examine the relationship of the share of both "high priced" (Group 1) and "low priced" (Group 3) tickets and realized load factors during the period prior to departure, which we define as the last 7 days.

In the language of our general specification, the outcome variable is the fraction of total seats sold that is comprised of Group 1 and Group 3 tickets sold in the last 7 days before departure. Table 5 presents the results. In our baseline set of flights that are "Empty/Empty", 21% of total tickets are Group 1 tickets sold in the last 7 days, and 14% of total tickets are Group 3 tickets sold in the last 7 days. (The other 65% of tickets are sold more than 7 days in advance or are Group 2 tickets sold in the last 7 days).

Column 1 of the table tests Dana's prediction that when demand (realized load factor) is high after controlling for expected load factor, the proportion of high priced (Group 1) seats sold in the last 7 days should also be high. The results do not support Dana's prediction. The only coefficient for realized load factor that is significant, Quartile 2, is of the wrong sign and nevertheless small in magnitude. Column 2 adds carrier-route fixed effects and this has no meaningful effect on the results.

The results for Group 3 tickets, in contrast, do offer support for Dana's predictions. Column 3 of Table 5 shows the relationship between the share of Group 3 tickets sold in the last 7 days prior to departure and realized load factor after controlling for expected load factor. The results show that when realized load factors are above the median, in either the third or fourth quartile, the share of Group 3 tickets sold during the period before departure declines. Relative to the baseline, tickets on flights in the highest quartile of realized load factor consist of 2% fewer Group 3 tickets, which represents a fourteen percent decline in the relative share of Group 3 tickets. These results offer modest evidence that flights with higher realized load factors have a smaller share of low-priced tickets.[18]

A complete treatment of how shares of tickets respond to changes in realized load factors requires an investigation of how these shares respond when realized load factor is high *during the booking process.* Such an approach is required because even if a flight is experiencing low bookings, say at seven days prior to departure, airlines may offer more cheaper discounted seats so that realized load factor at departure masks other changes that occurred during the process.

To investigate this issue, we conduct an analysis similar to that provided in Table 5, but where we measure realized load factor *at seven days prior to departure* (as opposed to using realized load factor at departure as we do above). Thus we test the

---

[18] Given the finding of a negative relationship between realized load factor and Group 3 share yet no relationship with Group 1 share, these results suggest a shift in share from Group 3 to Group 2. This is broadly consistent with the Dana model.

hypothesis that if we control for the realized load factor a week before departure, then seats sold in the last seven days are more likely to be more expensive (Group 1) tickets, as predicted by the Dana model (Prediction 4). We use seven days as a cutoff point because more than half of tickets are sold more than seven days before departure but over 40% of tickets remain to be sold. Further, the airline literature on yield management indicates that the number of tickets allocated to particular bins is changed in the week before departure if flights are significantly running ahead or behind in realized bookings.[19] Accordingly, we conduct an analysis like that in Table 5, but where realized load factor is measured at seven days prior to departure and we analyze the allocation of seats that are sold in the last seven days.

Table 6 presents the results. Beginning with "low price" Group 3 tickets in columns 3 and 4, the results show that when realized load factor is high seven days prior to departure, there is a notable drop in the share of Group 3 tickets sold in the last seven days of ticket sales. Column 4 reports a similar regression with carrier-route fixed effects and shows similar results. Conditional on the expected quartile of load factor, if a flight is realized to be in the highest quartile of load factor as of seven days before departure, then the fraction of tickets sold in the last seven days decreases by nearly five percentage points. When compared to a baseline that the unconditional mean share of tickets sold in the last seven days that are Group 3 is 28%, this represents an eighteen percent decrease. This suggests that if a flight is experiencing a high number of bookings as of a week before departure, the fraction of tickets sold in the last week that are "low price" is notably smaller. However, it is worth emphasizing that even flights that are both expected and realized to be high demand a week before departure will have a substantial number of Group 3 tickets sold in the last week.

Columns 1 and 2 show corresponding results for the shares of tickets sold in the last seven days that are "high price" Group 1 tickets. We find no statistically significant increase in the share of Group 1 tickets. These results viewed in conjunction with the Group 3 results above suggest that the reallocation involves shifting customers from

[19] In particular, two separate departments are involved in ticket sales. A pricing department sets prices on a carrier-route before booking begins by setting the fare that is attached to each combination of ticket characteristics and restrictions, which is sometimes referred to as a "fare bucket". Then, the yield management department allocates seats in each bucket at the beginning of booking, and may reallocate seats as booking proceeds. For a description of yield management, see Phillips (2005).

Group 3 "low price" tickets to Group 2 "medium price" tickets. This is suggestive evidence that on flights with high bookings a week before departure, passengers end up purchasing somewhat higher-priced tickets, either because the airlines have reallocated Group 3 tickets to higher priced groups or because the Group 3 tickets have sold out.

Regardless of the mechanism at play, the implication for our hypothesis testing is clear. Dana's model makes comparative static predictions that the share of high (low) priced tickets should rise (fall) on flights with higher realized load factors after controlling for the expected load factor. We find some evidence that is consistent with this prediction – the share of low-priced Group 3 tickets declines when comparing flights realized to be "Empty" to those realized to be "Full". Moreover, this relative decline is robust to analyzing all tickets sold (yielding a fourteen percent decline) and only tickets sold in the last seven days (yielding an eighteen percent decline).

4.D. Scarcity and the Dispersion of Fares

The tests above evaluate the impact of scarcity pricing on the allocation of tickets. A distinctly different question, however, is whether or not fare *dispersion* changes in response to changes in expected or realized load factors. Gale and Holmes theorize that peak flights are comprised of non-discounted fare passengers while off-peak flights consist of a mix of passengers buying both discounted and non-discounted fares. As a result, peak flights have less price dispersion (Prediction 7). Dana predicts that there is more fare dispersion on flights with high realized load factors, conditional on the expected load factor, because the low-priced tickets sell out before high-priced tickets offered on any flight (Prediction 2).

To test for these effects we use the empirical setup described above to investigate whether Gini coefficients are influenced by scarcity. We begin by using the allocation of flights into quartiles of expected and realized load factors as illustrated in Figure 2 at the carrier-route level. We then calculate Gini coefficients within these cells at the carrier-route level to determine the fare dispersion on flights with similar expected and realized load factors. These individual Gini coefficients at the carrier-route-cell level become the units of observation for the regressions. In all empirical specifications, we include carrier-route fixed effects to allow for differences in the dispersion of fares that are systematic across the carrier-route. The empirical tests investigate whether the variation in dispersion within a carrier-route is explained by expected and/or realized load factor, as predicted by the scarcity pricing theories.

The Gini coefficient has a convenient property for interpretation – the expected price difference of two randomly selected tickets divided by the mean fare equals twice the Gini coefficient. The average Gini coefficient across the carrier-routes for all tickets sold is 0.274. Therefore among the carriers and routes that we study, two randomly selected tickets on the same route and the same airline are expected to vary by 55% of the mean fare on that carrier-route.

Results are shown in Table 7. In the first column, we find only weak support for the Gale and Holmes prediction that peak flights have lower fare dispersion. Relative to flights expected to be Empty, flights expected to be in the second quartile of expected load factor have a Gini coefficient that is lower by -0.0075, which lowers the Gini coefficient by just under three percent. This is statistically different from the Gini in Empty flights, however the magnitude of the difference is economically quite small. For flights in higher quartiles of expected load factor, the Gini is not statistically different from the Gini on Empty flights.

In the second column, we test the prediction from Dana and find no support. Conditional on expected load factor, flights with higher realized load factor at departure have no more dispersion than those with low load factor. In fact, the point estimates, while not statistically significant, have the wrong sign.

As a robustness test, we test the Dana model using only tickets sold in the last seven days before departure. Suppose that yield management personnel adjust pricing as the departure date approaches in response to the number of seats that have already been sold on a flight. One might expect that the dispersion of fares sold in the last few days would change in response to this adjustment. The fare dispersion of tickets sold in the last seven days is slightly smaller than dispersion of overall fares—the mean Gini is 0.222 in the last seven days as compared to 0.274 overall. In column 3, we regress the Gini coefficient of tickets sold in the last seven days on the expected load factor and the realized load factor as of seven days before departure. We find that flights that are "running full" as of a week before departure do not have any more or less dispersion of fares sold in the last week.

This analysis of dispersion yields two important conclusions. First, there is a substantial amount of within carrier-route dispersion; as we discuss in the Introduction, this dispersion is larger than the dispersion across routes. Second, the amount of dispersion within a carrier-route does not vary in economically meaningful ways with either the expected or realized demand of flights. This suggests that models of scarcity pricing have virtually no ability to explain the large amount of differences in fares paid by passengers travelling on the same route and the same airline.

4.E. Scarcity and Average Pricing

In this section, we test the final two predictions -- the relationship between the level of fares and load factors. Gale and Holmes predict that flights with higher expected load factors (peak flights) have higher average fares than off-peak flights (Prediction 6). Dana predicts that the mean fare of transacted tickets is higher on flights with higher realized load factor, after conditioning on expected load factor (Prediction 1).

We test these predictions by specifying a variant of the general empirical specification (equation (1)) where the dependent variable is a metric of fare levels. In particular, first we calculate the average fare in each cell of Figure 2 for each carrier-route. Then for each carrier-route, we calculate the percentage difference of each cell from the Empty/Empty cell of that carrier-route, and these percentages serve as the dependent variable. Thus by construction, the dependent variable is 0% for the Empty/Empty cell, and the other cells measure the percentage by which that cell's mean transacted fare differs from the mean fare on Empty/Empty flights. As in the regressions above, an observation is a carrier-route-cell. The empirical specification tests whether the percentage deviation in mean fare from the baseline set of Empty/Empty flights is associated with higher levels of expected or realized load factor.

At first glance, this empirical model might appear to suffer from standard endogeneity concerns. After all, the specification is essentially a regression of "price" on "quantity". However, one should recall that the theoretical models offer clear comparative static predictions of the relationship between mean fares and equilibrium load factors. In the case of Gale and Holmes, flights with more tickets sold in expectation (i.e. "peak flights") will have higher average transacted fares. In the Dana model, after controlling for expected load factor, flights with a higher realization of tickets sold also have higher average fares on observed transactions.

Table 8 reports results. Column 1 tests the Gale and Holmes prediction that on-peak flights have higher priced transacted fares. We find that flights in higher quartiles of expected load factors have higher priced tickets. Relative to flights in the bottom quartile of expected load factor, flights in the second, third, and fourth quartiles have average fares that are 3%, 7%, and 9% higher, respectively.

The next three columns test Prediction 1 from the Dana model. Column 2 studies the effect of a higher realized load factor on the fares of all tickets sold, regardless of when those tickets are sold in the booking process. We find that relative to tickets on Empty/Empty flights, the fares of tickets that are realized to be in the top quartile of load factor are only slightly higher by 2%. This offers only modest support for Prediction 1 – the relationship is present only for "Full" flights and the magnitude is relatively small.

The specifications above test the literal interpretations of the theoretical pricing models. However, one could imagine pricing mechanisms that do not strictly follow the assumptions of the two theoretical models. In particular, both models envision a setting where airlines pre-commit to prices and do not adjust fares during the booking process in response to the contemporaneous load factor. In reality, airlines may use contemporaneous load factors to dynamically revise fares as the departure date approaches and demand shocks are observed. And as we discuss above, there is evidence from the yield management literature that this occurs in practice. Although fully developing such a model is beyond the scope of this paper, we can test for evidence that airlines dynamically adjust fares in a way that reflects scarcity pricing.

Therefore, we investigate a prediction that is inspired by the insights of the Dana model. We test if flights with the same expected load factor but a larger than expected number of tickets sold as of *X* days before departure, sell higher priced tickets in the last *X* days. One would expect a positive relationship regardless of whether airlines dynamically adjust fares or airlines keep initial fare buckets the same. If airlines do not dynamically adjust fares, the Dana model predicts a positive relationship, as we describe in section 2. If airlines dynamically raise fares when a flight is "running full" X days before departure, then the fares of tickets sold in the last X days should be higher.

Columns 3 and 4 of Table 8 show results for the tickets sold in the last fourteen and seven days, respectively. We find evidence that fares are notably higher in the last days before departure for flights with high contemporary load factors. Controlling for expected load factor, flights in higher quartiles of contemporary load factor have tickets that are priced higher. As compared to flights in the lowest quartile of contemporary load factor (i.e. those "running empty"), flights in the second, third, and fourth quartiles have tickets sold at fares that are 2%, 7%, and 13% higher. This phenomenon exists when analyzing both tickets sold in the last fourteen and in the last seven days.

5. Discussion and Conclusions

This paper provides rich evidence on how fares are related to the scarcity of seats on the flights of an individual itinerary. This paper tests implications of several of the leading models of scarcity-based pricing. Our results suggests that pricing models based upon scarcity provide insights that help explain some but not all dimensions of differences in the fares that passengers pay who travel on the same carrier and route.

First, we show that a great deal of the differences in fares is associated with basic institutional features of tickets such a refundability, travel, and stay restrictions. And it is noteworthy that the strong relationship between these ticket restrictions and fares is present *even on flights* that have little chance ex ante of selling out and/or do not sell out ex post.

Surprisingly, scarcity does not explain variation in the fare dispersion across different flights on the same carrier and route. In particular, the dispersion on flights that are expected and/or realized to be full is virtually the same as the dispersion on flights that have little chance of selling out and/or do not sell out.

However, the models of scarcity pricing explain some of the variation in average fares paid across flights. On-peak flights sell higher priced tickets – flights in the highest quartile of expected demand have fares that are 9% higher than lowest quartile. In addition, flights that are "running full" at 7 and 14 days before departure sell higher priced tickets in the days just before departure.

Regarding how tickets are allocated across flights, we find no strong evidence that discounted, advance purchase tickets are sold disproportionately on off-peak flights. However, we do find evidence that flights with unexpectedly high load factors observe a lower share of low-priced/highly restricted tickets.

Our findings provide the foundation for further empirical investigation on the nature of airline pricing. Future research could explore how ticket characteristics are key drivers of fares. Our finding that ticket characteristics are strongly associated with fares, even on low demand flights, is consistent with a variety of models of second-degree price discrimination, including work in the yield management literature and Dana's (1998) analysis of advance-purchase discounts. Such models have varying implications about the choice of capacity and the efficient use of that capacity. The role of ticketing restrictions can be explored in future work using our information on ticket characteristics.

References

Alderighi, Marco, Marcella Nicolini, and Claudio A. Piga, "Combined Effects of Load Factors and Booking Times on Fares: Insights from the Yield Management of a Low-Cost Airline", *Review of Economics and Statistics*, forthcoming.

Berry, Steven and Panle Jia, "Tracing the Woes: An Empirical Analysis of the Airline Industry", *American Economic Journal: Microeconomics*, 2(3): 1-43, 2010.

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

Borenstein, Severin. "U.S. Domestic Airline Pricing, 1995-2004", Berkeley, CPC Working Paper No. CPC05-48, January 2005.

Borenstein, Severin and Nancy Rose. "Competition and Price Dispersion in the U.S. Airline Industry," *Journal of Political Economy*, 102(4):653-683, August 1994.

Borenstein, Severin and Nancy Rose. "Bankruptcy and Pricing Behavior in U.S. Airline Markets," *American Economic Review*, 85(2):397-402, May 1995.

Borenstein, Severin and Nancy Rose. "How Airline Markets Work…Or Do They? Regulatory Reform in the Airline Industry", Economic Regulation and Its Reform: What Have We Learned?, University of Chicago Press, Nancy Rose ed., 2014.

Busse, Meghan. "Firm Financial Condition and Airline Price Wars," *RAND Journal of Economics*, 33(2): 298-318, Summer 2002.

Ciliberto, Federico and Carola Schenone. 2012. "Are the Bankrupt Skies the Friendliest?", *Journal of Corporate Finance*, 18: 1217-1231.

Dana, James D., Jr. 1998. "Advance-Purchase Discounts and Price Discrimination in Competitive Markets." *Journal of Political Economy*, 106 (2): 395-422.

Dana, James D., Jr. 1999a. "Using Yield Management to Shift Demand when the Peak Time is Unknown." *RAND Journal of Economics*, 30 (Autumn): 456-74.

Dana, James D., Jr. 1999b. "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure." *RAND Journal of Economics,* 30 (Winter): 632-60.

Dana, James D. and Eugene Orlov. 2014. "Internet Penetration and Capacity Utilization in the US Airline Industry", *American Economic Journal: Microeconomics*, 6(4): 106-137.

Dobson, Paul W. and Claudio A. Piga. 2013. "The Impact of Mergers on Fare Structure: Evidence from European Low-Cost Carriers", *Economic Inquiry*, 51(2): 1196-1217.

Eden, Benjamin. 1990. "Marginal Cost Pricing When Spot Markets are Complete." *Journal of Political Economy*, 98 (6) 1293-306.

Escobari, Diego. 2009. "Systematic Peak-load Pricing, Congestion Premia and Demand Diverting: Empirical Evidence". *Economics Letters*, 103(1): 59-61.

Escobari, Diego. 2012. "Dynamic Pricing, Advance Sales, and Aggregate Demand Learning in Airlines". *Journal of Industrial Economics*, 60(4): 697-724.

Escobari, Diego. 2014. "Estimating Dynamic Demand for Airlines", *Economics Letters*, 124(1): 26-29.

Escobari, Diego and Gan, Li. 2007. "Price Dispersion under Costly Capacity and Uncertain Demand". NBER Working Paper No. 13075, May 2007.

Forbes, Silke Januszewski, "The Effect of Air Traffic Delays on Airline Prices", *International Journal of Industrial Organization*, 26(5): 1218-1232, September 2008.

Gaggero, Alberto A. and Claudio A. Piga. 2011. "Airline Market Power and Intertemporal Price Discrimination", *Journal of Industrial Economics*, LIX(4): 552-577.

Gale, Ian, and Thomas Holmes. 1992. "The Efficiency of Advance-Purchase Discounts in the Presence of Aggregate Demand Uncertainty." *International Journal of Industrial Organization* 10 (3): 413-37.

Gale, Ian and Thomas Holmes. 1993. "Advance-Purchase Discounts and Monopoly Allocation of Capacity." *American Economic Review* 83 (March): 135-46.

Gallego, Guillermo and Garrett van Ryzin 1994. "Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons", *Management Science*, 40(8): 999-1020.

Gerardi, Kristopher S. and Adam Hale Shapiro. 2009. "Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry", *Journal of Political Economy*, 117(1): 1-37, February 2009.

Goolsbee, Austan and Chad Syverson. 2008. "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." *Quarterly Journal of Economics*, 123(4):1611-1633.

Lazarev, John. 2013. "The Welfare Effects of Intertemporal Price Discrimination: An Empirical Analysis of Airline Pricing in U.S. Monopoly Markets", mimeo, New York University.

Lederman, Mara. 2008. "Are Frequent-Flyer Programs a Cause of the "Hub Premium"?, *Journal of Economics and Management Strategy*, 17(1): 35-66.

McAfee, R. Preston, and Vera L. te Velde. 2006. "Dynamic Pricing in the Airline Industry." In Handbook on Economics and Information Systems, vol. 1, edited by T.J. Hendershott. New York: Elsevier Science.

Phillips, Robert L. Pricing and Revenue Optimization, Stanford University Press, 2005.

Prescott, Edward C. 1975. "Efficiency of the Natural Rate." *Journal of Political Economy* 83 (6): 1229-36.

Sengupta, Anirban and Steven N. Wiggins. 2014. "Airline Pricing, Price Dispersion, and Ticket Characteristics On and Off the Internet." *American Economic Journal: Economic Policy*, 6(1): 272-307.

Smith. B, Rao. B and Ratliff. R. 2001. "E-Commerce and Operations Research in Airline Planning, Marketing, and Distribution," *Interfaces*, 31(2): 37-55.

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

Stahl, D. 1996. "Oligopolistic Pricing with Heterogeneous Search", October, *International Journal of Industrial Organization*, 14, 1996: 243–268.

Stavins, Joanna. 2001. "Price Discrimination in the Airline Market: The Effect of Market Concentration." *Review of Economics and Statistics*, 83 (1): 200-02.

Williams, Kevin R. 2013. "Dynamic Airline Pricing and Seat Availability", mimeo, Yale.

Online Appendix: Data

Transactions Data

We study itineraries for travel in 2004Q4 that were purchased between June and December 2004 through the Computer Reservation System (CRS) that provided us with the data. Although we do not have data on transactions occurring prior to June (which means we miss transactions occurring 4 months before our first day of October 1, 2004), we do not expect this to substantively affect our results.

We exclude itineraries involving any international travel, more than four coupons, open jaws and circular trips, or more than one carrier. Also, we exclude itineraries with a zero fare.

We calculate a measure of flight level load factor using the tickets we observe and the CRS's share of tickets sold on a city-pair. The CRS share is calculated by finding the fraction of total coupons for non-stop travel between two cities (the "T-100 Domestic Segment" data from the Bureau of Transportation Statistics) that we observe in our transaction data. We compute these "CRS shares" at the route-carrier level.

Procedure to Merge Transaction Data to Posted Fare Data

We used the following procedure to match transactions from the CRS providing us with transaction data to posted fares from the CRS that provided archived fares.

In the first step, we matched a ticket from the transaction data to a posted fare using carrier, date of departure (but not return), booking class, and price.[20] In this first step, we included any fares matching within 10%.

After this first step, the resulting dataset included multiple matching posted fares for some individual transactions. This primarily included multiple matching fares with different combinations of advance purchase requirements and travel restrictions. Because our transaction data include no additional information to facilitate matching, we were

[20] In this first matching step, we only require fares to match within 10%. In a later step, we require fares to match much closer. In addition, we matched a transaction's date of departure to a 7-day window of days of departure in the posted fare data, and later use the match in which the dates of departure are closest.

required to make additional assumptions. In the second step of the matching procedure, we eliminate multiple matches on advance purchase. We assume that the ticket was purchased with the most restrictive advance purchase requirement for which it qualified.[21]

For any transactions that still matched multiple posted fares, we adopted a third matching step. Prices were required to match within a 2 percent range.[22] Any remaining multiple matches were then screened to meet travel restrictions that involve travel on specified days of the week. For example, some posted fares required travel on a Tuesday, Wednesday or Thursday. Using the ticket's date of departure, we eliminated any multiple matches that did not satisfy the posted travel restriction. For any additional transactions with multiple matches, we assumed that any ticket meeting a travel restriction had that travel restriction. For example, a ticket matching fares with and without a travel restriction was assumed to have that travel restriction.

The final step includes the verification of minimum and maximum stay restrictions. For the minimum and maximum stay restrictions collected from the travel agent, some restrictions were explicitly given (namely 1 day, 2 days etc.). However, other posted fares were indicated to include a travel restriction but the restriction was not specifically named on the travel agent's CRS screen that we accessed. For the matches where the minimum and maximum days of stay restriction were given, we verified that the actual transactions met the specific requirements. In case of multiple matches (which comprise less than 1%), if two tickets had the same characteristics but one required a 1 day minimum stay while the other did not, and the transaction involved a 2 day stay, we match the posted fare with a 1 day minimum stay.

[21] For example, suppose a ticket was purchased 16 days before departure. If the first step matched both a 14 day and a 7 day advance purchase requirement, we match the transaction with the posted fare that required a 14 day advance purchase.

[22] We should note that the local travel agent used a different CRS than our transaction data. Since July 2004, CRSs were not required to post identical fares.

Figure 1

Comparing the Kernel Densities of Matched and Unmatched Transactions
All Carriers and All Routes

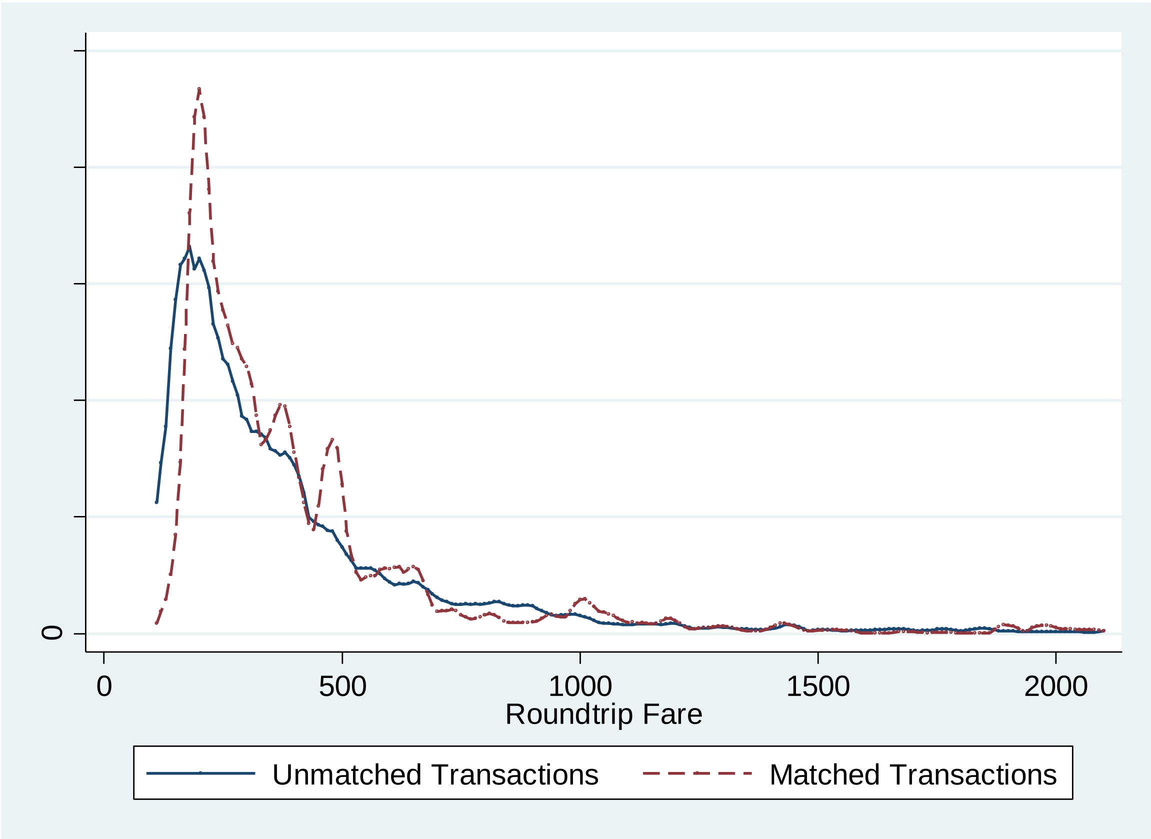

Figure 2

Dividing Sample by Expected and Realized Load Factors

	Expected Load Factors			
	Full	Medium-Full	Medium-Empty	Empty
Realized Load Factors				
Full				
Medium-Full				
Medium-Empty				
Empty				

This table illustrates how flights are divided to test comparative static predictions about the characteristics of tickets sold on flights that are unusually full on peak flights and unusually empty on off-peak flights. We divide flights (flight-departure date) into quartiles based upon expected load factor and then actual load factor. We create the categories so there are approximately the same number of tickets in each cell. A complete description of the methodology is included in the text.

Table 1: Routes Included in Analysis

The analysis includes all carriers flying on any of these routes, where the routes are large routes for the six carriers below.

American					
	LAS-DFW	LAX-JFK	PHX-DFW	DFW-DEN	ORD-STL
	LAX-DFW	ORD-LGA	LAX-ORD	ORD-DFW	DFW-MCO
	SJU-MIA	STL-DFW	DFW-SNA	LGA-MIA	MIA-JFK
Delta					
	DFW-ATL	LAS-ATL	ATL-MIA	ATL-PHL	EWR-ATL
	MCO-ATL	LGA-ATL	TPA-ATL	ATL-FLL	BOS-ATL
	LAX-ATL	CVG-ATL	CVG-LGA	FLL-BDL	LAX-TPA
United					
	LAX-DEN	LAS-ORD	IAD-ORD	LAS-DEN	SEA-ORD
	LAX-ORD	DEN-ORD	ORD-SFO	SFO-LAX	ORD-LGA
	SFO-SAN	IAD-SFO	OAK-DEN	ONT-DEN	PDX-SFO
Continental					
	LAX-EWR	DEN-IAH	ORD-IAH	ATL-EWR	IAH-DFW
	EWR-MCO	FLL-EWR	LAS-EWR	BOS-EWR	SFO-EWR
	IAH-LAX	EWR-IAH	MSY-IAH	IAH-LAS	IAH-MCO
Northwest					
	MSP-PHX	MSP-LAS	DEN-MSP	DTW-LAS	PHX-DTW
	LGA-DTW	MCO-DTW	LAX-DTW	MSP-MCO	MKE-MSP
	DTW-MSP	LAX-MSP	SEA-MSP	MSP-SFO	BOS-DTW
USAir					
	PHL-MCO	FLL-PHL	BOS-DCA	BOS-LGA	ORD-PHL
	PHL-BOS	LGA-DCA	PHL-TPA	LAS-PHL	RDU-PHL
	MCO-CLT	CLT-PHL	LGA-CLT	CLT-BOS	PIT-PHL

Notes: These routes are large representative routes for each of the six carriers. Airport codes: ATL=Atlanta, BDL=Hartford, BOS=Boston, CLT=Charlotte, CVG=Cincinnati, DCA=Washington-Reagan, DEN=Denver, DFW=Dallas-FtWorth, DTW=Detroit, EWR=Newark, FLL=Fort Lauderdale, IAD=Washington-Dulles, IAH=Houston, JFK=NY-JFK, LAS=Las Vegas, LAX=Los Angeles Intl, LGA=NY-La Guardia, MCO=Orlando, MIA=Miami, MKE=Milwaukee, MSP=Minneapolis-St Paul, MSY=New Orleans, OAK=Oakland, ONT=Ontario, ORD=Chicago-O'Hare, PDX=Portland, PHL=Philadelphia, PHX=Phoenix, PIT=Pittsburgh, RDU=Raleigh-Durham, SAN=San Diego, SEA=Seattle, SFO=San Francisco, SJU=San Juan, SNA=Orange County, STL=St. Louis, TPA=Tampa.

Table 2: Summary Statistics

Panel A: Sample Means

Variable	All Transactions	Matched Transactions
Fare (for roundtrip)	$415.48	$424.25
Refundable	--	0.26
Some Travel Restriction (e.g. DOW)	--	0.38
Minimum Stay Restriction	--	0.20
Maximum Stay Restriction	--	0.15
Purchased 0-3 Days in Advance	0.28	0.31
Purchased 4-6 Days in Advance	0.14	0.14
Purchased 7-13 Days in Advance	0.20	0.20
Purchased 14-21 Days in Advance	0.14	0.14
Purchased > 21 Days in Advance	0.24	0.21
Roundtrip Itinerary	0.66	0.65
American	0.30	0.28
Delta	0.16	0.15
United	0.15	0.15
Continental	0.16	0.18
Northwest	0.07	0.08
USAir	0.17	0.15
Monday Departure	0.19	0.20
Tuesday Departure	0.16	0.18
Wednesday Departure	0.16	0.17
Thursday Departure	0.15	0.16
Friday Departure	0.16	0.13
Saturday Departure	0.07	0.06
Sunday Departure	0.11	0.11
N	612,903	221,895

Note: This panel contains summary statistics for itineraries to travel in 2004Q4 on American, Delta, United, Northwest, Continental and USAir on the routes in our sample. The first column includes all transactions through the CRS that gave us transaction data (excluding first class tickets and itineraries involving more than four coupons, as discussed in the Data section). The second column includes only transactions we were able to match with ticket characteristics from the other CRS's archive.

Panel B: Mean Gini Coefficient By Cell Across All Carrier-Routes

		Expected Load Factor			
		Full	Medium-Full	Medium-Empty	Empty
Realized Load Factor	Full	0.276	0.274	0.274	0.274
	Medium-Full	0.272	0.275	0.268	0.277
	Medium-Empty	0.276	0.276	0.273	0.277
	Empty	0.270	0.275	0.267	0.284

Note: This panel contains the mean Gini coefficient across all carrier-routes for a given cell of the matrix of quartiles of expected and realized load factors. To calculate these figures, we first calculate for each carrier-route the Gini coefficient of all tickets sold on flights in each cell of the matrix (see description of calculating expected and realized load factor in the Data section). This will yield 16 Gini coefficients for each carrier-route -- one for each cell of the matrix. This table reports the mean Gini across all the carrier-routes for the respective cell of the matrix.

Table 3: Relationship Between Fares and Ticket Restrictions

Dependent Variable: log(fare)

	(1)	(2)	(3)
	All Transactions	Only itineraries on "Empty/Empty" Flights	All Transactions
Refundable	0.56*	0.56*	
	(0.01)	(0.03)	
Roundtrip Itinerary	-0.20*	-0.18*	
	(0.01)	(0.01)	
Day of Travel Restriction	-0.33*	-0.31*	
	(0.00)	(0.01)	
Length of Stay Restriction	-0.14*	-0.16*	
	(0.01)	(0.01)	
Group 1 (high-price, unrestricted)			0.51*
			(0.01)
Group 3 (low-price, highly restricted)			-0.45*
			(0.01)
Constant	6.53*	6.24*	6.46*
	(0.07)	(0.23)	(0.07)
# Observations	221,895	12,753	221,895
R^2	0.656	0.663	0.627

* significant at the 1% level

Note: The unit of observation is an itinerary. Models included fixed effects for carrier-route and the week of the year of travel. The model is estimated via least squares with robust standard errors clustered on the date of initial departure reported in parentheses. The R^2 of a model with only the fixed effects is 0.352. In column 2, "Empty/Empty" flights, as described in section 3.C, are those flights in the bottom quartile of each carrier-route's expected load factor and, among those flights, are in the bottom quartile of realized load factor. In column 3, the categorizing of tickets into Groups based upon ticket characteristics is described in section 4.A; the excluded category is Group 2, which are tickets with limited restrictions, i.e. tickets that are nonrefundble but do not carry day of travel or length of stay restrictions.

Table 4: Test of Gale/Holmes Prediction that Peak Flights Have Fewer Group 3 Tickets Sold in Advance

Dependent Variable: Share of all tickets sold that are Group 3 tickets Sold in Advance

	(1)	(2)	(3)	(4)
	> 21 Days in Advance		**> 14 Days in Advance**	
Expected Load Factor - 2nd Quartile	-0.0094	-0.0094	-0.0079	-0.0079
	(0.0052)	(0.0054)	(0.0057)	(0.0059)
Expected Load Factor - 3rd Quartile	-0.0084	-0.0084	-0.0072	-0.0072
	(0.0057)	(0.0059)	(0.0069)	(0.0071)
Expected Load Factor - 4th Quartile	-0.0091	-0.0091	-0.0089	-0.0089
	(0.0066)	(0.0068)	(0.0079)	(0.0081)
Constant	0.1621*	0.1516*	0.2453*	0.2187*
	(0.0113)	(0.0039)	(0.0149)	(0.0046)
Carrier-Route Fixed Effects	No	Yes	No	Yes
# Observations	1904	1904	1904	1904
R^2	0.0008	0.7108	0.0004	0.7761

* significant at the 1% level + significant at 5% level

Note: The unit of observation is a cell of the 4x4 matrix of expected and realized load factor in Figure 2 for each combination of carrier and route. The model is estimated via least squares with robust standard errors clustered on the carrier-route and reported in parentheses. Group 3 tickets are highly restricted tickets (nonrefundable with travel and/or stay restrictions) that have lower fares, as described in section 4.A.

Table 5: Test of Dana Prediction that Share of High Priced Tickets is Higher on Flights with Higher Realized Load Factor After Controlling for Expected Load Factor (Using ALL Transactions)

Dependent Variable: Share of ALL tickets that are sold in last 7 days before departure and belong to Group 1 (or Group 3)

	(1)	(2)	(3)	(4)
	Group 1		Group 3	
Realized Load Factor at Departure - 2nd Quartile	-0.0073+	-0.0073	-0.0036	-0.0036
	(0.0036)	(0.0037)	(0.0037)	(0.0039)
Realized Load Factor at Departure - 3rd Quartile	-0.0027	-0.0027	-0.0102+	-0.0102+
	(0.0042)	(0.0044)	(0.0044)	(0.0046)
Realized Load Factor at Departure - 4th Quartile	-0.0053	-0.0053	-0.0202*	-0.0202*
	(0.0053)	(0.0054)	(0.0065)	(0.0067)
Expected Load Factor - 2nd Quartile	0.0033	0.0033	-0.0033	-0.0033
	(0.0052)	(0.0054)	(0.0059)	(0.0061)
Expected Load Factor - 3rd Quartile	0.0075	0.0075	-0.0017	-0.0017
	(0.0056)	(0.0058)	(0.0063)	(0.0065)
Expected Load Factor - 4th Quartile	0.0105	0.0105	-0.0105	-0.0105
	(0.0063)	(0.0065)	(0.0065)	(0.0067)
Expected 1st or Realized 1st But Not Both	-0.002	-0.002	0.006	0.006
	(0.0032)	(0.0033)	(0.0031)	(0.0032)
Constant	0.2081*	0.1927*	0.1420*	0.0814*
	(0.0246)	(0.0060)	(0.0128)	(0.0067)
Carrier-Route Fixed Effects	No	Yes	No	Yes
# Observations	1904	1904	1904	1904
R^2	0.0004	0.9489	0.0056	0.7675

* significant at the 1% level + significant at 5% level

Note: The unit of observation is a cell of the 4x4 matrix of expected and realized load factor in Figure 2 for each combination of carrier and route. These regressions use all transactions. The model is estimated via least squares with robust standard errors clustered on the carrier-route and reported in parentheses. Group 1 tickets are unrestricted tickets (refundable) that have higher fares, and Group 3 tickets are highly restricted tickets (nonrefundable with travel and/or stay restrictions) that have lower fares, as described in section 4.A.

Table 6: Test of Dana Prediction that Share of High Priced Tickets is Higher on Flights with Higher Realized Load Factor After Controlling for Expected Load Factor (Using Only Transactions Occuring Within 7 Days of Departure)

Dependent Variable: Share of tickets sold in last 7 days before departure that belong to Group 1 (or Group 3)

	(1)	(2)	(3)	(4)
	Group 1		**Group 3**	
Realized Load Factor *at 7 Days* - 2nd Quartile	-0.0038	-0.0039	-0.0119	-0.011
	(0.0055)	(0.0057)	(0.0060)	(0.0063)
Realized Load Factor *at 7 Days* - 3rd Quartile	0.0128	0.0127	-0.0283*	-0.0273*
	(0.0074)	(0.0077)	(0.0078)	(0.0081)
Realized Load Factor *at 7 Days* - 4th Quartile	0.0091	0.009	-0.0472*	-0.0438*
	(0.0085)	(0.0089)	(0.0098)	(0.0101)
Expected Load Factor - 2nd Quartile	-0.0002	-0.0001	0.0011	-0.0007
	(0.0073)	(0.0077)	(0.0093)	(0.0096)
Expected Load Factor - 3rd Quartile	0.0102	0.0103	-0.0101	-0.0118
	(0.0071)	(0.0073)	(0.0096)	(0.0099)
Expected Load Factor - 4th Quartile	0.0155	0.0155	-0.0209	-0.0226
	(0.0087)	(0.0090)	(0.0123)	(0.0127)
Expected 1st or Realized 1st But Not Both	-0.0017	-0.0017	-0.0058	-0.0041
	(0.0060)	(0.0062)	(0.0059)	(0.0059)
Constant	0.3410*	-0.0522*	0.3084*	0.8255*
	(0.0332)	(0.0080)	(0.0237)	(0.0096)
Carrier-Route Fixed Effects	No	Yes	No	Yes
# Observations	1904	1904	1904	1904
R^2	0.0007	0.9285	0.0054	0.7824

* significant at the 1% level + significant at 5% level

Note: The unit of observation is a cell of the 4x4 matrix of expected and realized load factor in Figure 2 for each combination of carrier and route. In contrast to Table 5, these regressions only use transactions that occur in the seven days before departure and measure realized load factor as of seven days before departure. The model is estimated via least squares with robust standard errors clustered on the carrier-route and reported in parentheses. Group 1 tickets are unrestricted tickets (refundable) that have higher fares, and Group 3 tickets are highly restricted tickets (nonrefundable with travel and/or stay restrictions) that have lower fares, as described in section 4.A.

Table 7: Test of Relationship Between Fare Dispersion and Load Factor

Dependent Variable: Gini coefficient of all transactions in carrier-route-cell

	(1)	(2)	(3)
	Gale/Holmes	Dana	
	All Transactions	All Transactions	Transactions in Last 7 Days
Expected Load Factor - 2nd Quartile	-0.0075*	-0.0102*	-0.0021
	(0.0027)	(0.0028)	(0.0036)
Expected Load Factor - 3rd Quartile	-0.0027	-0.0055	0.0025
	(0.0036)	(0.0036)	(0.0040)
Expected Load Factor - 4th Quartile	-0.0046	-0.0073	0.0004
	(0.0042)	(0.0043)	(0.0040)
Realized Load Factor at Departure - 2nd Quartile		-0.0013	
		(0.0020)	
Realized Load Factor at Departure - 3rd Quartile		-0.0041	
		(0.0026)	
Realized Load Factor at Departure - 4th Quartile		-0.0022	
		(0.0031)	
Realized Load Factor *at 7 Days* - 2nd Quartile			-0.0053
			(0.0031)
Realized Load Factor *at 7 Days* - 3rd Quartile			0.0025
			(0.0037)
Realized Load Factor *at 7 Days* - 4th Quartile			-0.004
			(0.0057)
Expected 1st or Realized 1st But Not Both		-0.0054*	-0.0094*
		(0.0017)	(0.0026)
Constant	0.3914*	0.3973*	0.3973*
	(0.0023)	(0.0034)	(0.0034)
Carrier-Route Fixed Effects	Yes	Yes	Yes
Mean Gini Across Carrier-Route-Cells	0.274	0.274	0.222
# Observations	1904	1904	1903
R^2	0.8409	0.8416	0.700

* significant at the 1% level   + significant at 5% level

Note: The unit of observation is a cell of the 4x4 matrix of expected and realized load factor in Figure 2 for each combination of carrier and route. The model is estimated via least squares with robust standard errors clustered on the carrier-route and reported in parentheses. Column 1 and 2 use all transactions while Column 3 uses only transactions that occur within seven days of departure.

Table 8: Test of Relationship Between Average Fares Paid and Load Factor

Dependent Variable: Percentage Difference of Average Flight Fare of carrier-route-cell from Carrier-Route Mean of Empty-Empty

	(1)	(2)	(3)	(4)
	Gale/Holmes	Dana		
	All Sales	All Sales	Sales in Last 14 Days	Sales in Last 7 Days
Expected Load Factor - 2nd Quartile	0.03*	0.02	0.02+	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
Expected Load Factor - 3rd Quartile	0.07*	0.06*	0.07*	0.05*
	(0.01)	(0.01)	(0.01)	(0.01)
Expected Load Factor - 4th Quartile	0.09*	0.08*	0.07*	0.06*
	(0.01)	(0.01)	(0.01)	(0.01)
Realized Load Factor at Departure - 2nd Quartile		-0.01+		
		(0.01)		
Realized Load Factor at Departure - 3rd Quartile		0.00		
		(0.01)		
Realized Load Factor at Departure - 4th Quartile		0.02+		
		(0.01)		
Realized Load Factor *at 14 Days* - 2nd Quartile			0.02*	
			(0.01)	
Realized Load Factor *at 14 Days* - 3rd Quartile			0.07*	
			(0.01)	
Realized Load Factor *at 14 Days* - 4th Quartile			0.13*	
			(0.01)	
Realized Load Factor *at 7 Days* - 2nd Quartile				0.02*
				(0.01)
Realized Load Factor *at 7 Days* - 3rd Quartile				0.07*
				(0.01)
Realized Load Factor *at 7 Days* - 4th Quartile				0.13*
				(0.02)
Expected 1st or Realized 1st But Not Both		-0.02*	-0.01+	-0.01
		(0.01)	(0.01)	(0.01)
Constant	-0.01	0.00	0.00	0.00
	(0.01)	(0.00)	(0.00)	(0.00)
# Observations	1904	1904	1904	1903
R^2	0.0413	0.0503	0.1075	0.0909

* significant at the 1% level + significant at 5% level

Note: The unit of observation is a cell of the 4x4 matrix of expected and realized load factor in Figure 2 for each combination of carrier and route. The model is estimated via least squares with robust standard errors clustered on the carrier-route and reported in parentheses. The dependent variable is the percentage difference of mean fare for each carrier-route-cell from the carrier-route's average fare on "Empty/Empty" flights. To construct the dependent variable, first we calculate the average fare in each cell of Figure 2 for each carrier-route. Then for each carrier-route, we calculate the percentage difference of each cell from the Empty/Empty cell of that carrier-route, and these percentages serve as the dependent variable. Thus by construction, the dependent variable is 0% for the Empty/Empty cell, and the other cells measure the percentage by which that cell's mean transacted fare differs from the mean fare on Empty/Empty flights