Working Papers

Federal Reserve Bank of Boston

No. 07-7

Does Competition Reduce Price Discrimination? New Evidence from the Airline Industry

Kristopher Gerardi and Adam Hale Shapiro

Abstract:
This paper analyzes the effects of market structure on price dispersion in the airline industry, using panel data from 1993 through 2006. The results found in this paper contrast with those of Borenstein and Rose (1994), who found that price dispersion increases with competition. We find that competition has a negative effect on price dispersion, in line with the traditional textbook treatment of price discrimination. Specifically, the effects of competition on price dispersion are most significant on routes that we identify as having consumers characterized by relatively heterogeneous elasticities of demand. On routes with a more homogeneous customer base, the effects of competition on price discrimination are largely insignificant. We conclude from these results that competition acts to erode the ability of a carrier to price discriminate, resulting in reduced overall price dispersion.

JEL Classifications: D43, L13, L93

Kristopher Gerardi and Adam Hale Shapiro are research associates at the Federal Reserve Bank of Boston and Ph.D. candidates at Boston University. Their email addresses are, respectively, kristopher.gerardi@bos.frb.org and adam.shapiro@bos.frb.org.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm.

The authors are grateful to Susanto Basu, Simon Gilchrist, Lorenz Goette, Bob King, Marc Rysman, and Joanna Stavins for helpful comments and suggestions.

This version: September 2007

1 Introduction

The relationship between price discrimination and market structure has been the focus of a great deal of economic research. Traditional textbooks argue that market power increases the firm's ability to sustain markups, and thus increases the firm's ability to implement price discrimination strategies.[1] It follows that competitive firms cannot price discriminate since they are price takers, while monopolists can price discriminate to the extent that there exists both heterogeneity in consumers' demand elasticities and a usable sorting mechanism to distinguish between consumer types. The textbook theory therefore predicts that *ceteris paribus*, more concentrated markets should be characterized by more price discrimination, since increased market power implies more discretion in setting prices. However, a number of theoretical studies have shown that this may not be the case. Theoretical papers by Borenstein (1985) and Holmes (1989) showed that the existence of brand loyalty in imperfectly competitive markets could create a *negative* relationship between concentration and price discrimination. As there has been no overarching theoretical model, the relationship between market structure and price discrimination becomes an empirical question.

The airline industry has been the focus of many price-discrimination studies because of the prominence of price discrimination in airlines' pricing strategies. Furthermore, the industry is rich with historical data. In the airline industry, price discrimination is used by firms to distinguish between relatively price-inelastic business travelers and relatively price-elastic leisure travelers. In a seminal empirical study of the airline industry, Borenstein and Rose (1994) found evidence of a *negative* relationship between price discrimination and concentration. The authors found that routes characterized by higher levels of competition exhibited more price dispersion. They attributed this result to airline pricing practices that are based on exploiting heterogeneity in customers' brand preference, rather than solely on heterogeneity in reservation prices for air travel. That is, business travelers remain loyal to an airline and ignore lower airfare offered by competitors. In contrast, leisure travelers, who have weaker preferences for specific brands than business travelers, are more apt to purchase lower airfare regardless of brand. Borenstein and Rose's findings spawned a new line in the airline literature that attempted to verify the existence of a negative relationship between concentration and price dispersion.[2]

In the decade or so since the Borenstein and Rose study, the U.S. domestic airline industry has experienced large changes in competition, demand, and cost. Large increases

[1] See, for example, Varian (1989).

[2] Empirical studies include Stavins (1996) and Hayes and Ross (1998). For an example of a theoretical study, see Gale (1993).

in the price of oil, beginning at the turn of the century, have placed severe upward pressure on airlines' input costs. The latest business cycle, characterized by the Information Technology (IT) boom in the late 1990s and the subsequent recession in the early 2000s, and further incited by the September 11th terrorist attacks, also has had a large impact on the demand for domestic air travel. As a result, a number of the large, traditional U.S. carriers, or "legacy carriers," have been forced to declare bankruptcy in recent years. At the same time a number of new airlines, with vastly different operating strategies have emerged. These aptly named, "low-cost carriers" (LCCs) have increased competition along many domestic routes, threatening the remaining market shares of the traditional legacy carriers. Although LCCs initially entered the market in pursuit of leisure travelers, the media have documented that their emergence has also benefited business travelers. For instance,

> ...even as the recovery has gained steam, it's clear that price-sensitive business travelers aren't returning to the days when they readily paid the highest fares. Instead, they're buying cheaper, restricted tickets or moving to the low-cost carriers. Last year (2003), the average fare paid by business travelers was 49-percent less than the typical business fare offered by airlines.
> -*Business Week*, May 24, 2004

Thus, it is quite conceivable that business travelers, as well as leisure travelers, have benefited from the lower airfare offered by these new competitors.

In light of these observations, we set out to re-examine Borenstein and Rose's findings by using more complete data on airline prices. We construct a 13-year panel of domestic airline ticket prices covering the period from 1993 to 2006. In analyzing the cross-section, we are able to replicate their results for most of the quarters in our sample. That is, we find a negative relationship between concentration and price dispersion when using their specification on the cross-section. However, when we control for time-invariant, route-specific effects using panel-data methods, we find that the estimated relationship between concentration and price dispersion becomes positive. We find that one possible source of the difference between the Borenstein and Rose cross-sectional estimates and the panel estimates is an omitted-variable bias and an invalid instrument, which we explore in further detail below.

In addition, this study finds that the effect of competition on price dispersion is stronger on routes that are identified as having heterogeneous customer bases, namely, relatively large numbers of both business and leisure travelers. Following an increase in competition on these routes, we find that prices in the upper part of the price distribution fall by a larger amount than do prices in the lower part. Furthermore, we find that this effect is strongest when the competitor is a low-cost or regional carrier. On routes with mostly leisure travelers, competition generally does not have a statistically significant effect on price dispersion. Our study therefore suggests that business travelers are more affected by an increase in competition than leisure travelers, and also that the entry by LCCs or regional carriers has a stronger competitive effect than does entry by legacy carriers.

Given our empirical findings, we conclude that the observed fall in price dispersion is due to the erosion of a carrier's ability to price discriminate, in line with the textbook theoretical explanation of sustaining markups in monopoly markets. In addition to theoretical implications, our analysis also has potentially important public-policy ramifications. The consequences for consumers of deregulation and privatization in a particular industry depend on the relationship between competition and pricing in that industry. As our study shows that airlines are better able to price discriminate in more concentrated markets, policies aimed at increasing the ease of entering markets will narrow the gap between the prices charged to business and leisure consumers.

In Section 2, we discuss the empirical and theoretical literature regarding the relationship between price discrimination and market structure. Section 3 contains a detailed discussion of the data and our method of partitioning flights into those that are likely to contain both business and leisure travelers and those that are likely to contain predominantly leisure travelers. Section 4 presents cross-sectional regressions that update and extend those of Borenstein and Rose (1994). Section 5 includes our own fixed-effects panel analysis. Section 6 discusses route-specific factors that may play an important role in the cross-section, and Section 7 concludes.

2 Literature Review

2.1 Sources of Price Dispersion in the Airline Industry

There are a number of reasons why firms in certain industries charge different prices for the same good. Perhaps the most apparent reason is to exploit customer heterogeneity in reservation prices. If customers for a product are heterogenous, and if a firm with market power can learn the reservation price of each customer, then profit maximization calls for the firm to evaluate each customer separately and set prices accordingly. This strategy, known as "first-degree price discrimination," requires large amounts of customer-specific information, which is nearly impossible to obtain in most industries. A more practical strategy is to target individuals based on observable characteristics, a practice known as "third-degree price discrimination." This type of price discrimination in the airline industry takes the form of advance-purchase requirements and other types of restrictions on ticket purchases. Many airlines offer fully refundable tickets at relatively high prices, while offering non-refundable tickets at much lower prices. Another method historically used to price discriminate is requiring Saturday-night stayovers. Since many consumers travel for business-related reasons and tend to place a high value on their time, they are more likely to purchase more expensive tickets without such restrictions. Advance-purchase requirements also allow firms to distinguish between consumer types. In this case, airlines exploit the tendency of price-inelastic consumers to purchase tickets close to the date of departure, when, for example, necessary, last-minute business trips or family emergencies occur. By making use of these techniques, airlines are able to separate price-sensitive travelers from price-insensitive travelers.

There are other ways besides price discrimination by which a firm can induce multiple prices for the same good. "Peak-load pricing" is one such pricing strategy in which airlines change prices to alleviate potential capacity constraints during times of predictably high demand. For example, on many routes there is much higher demand for weekday flights than for flights on the weekends, so airlines will accordingly set higher prices during the week.[3] Unfortunately, our data provide us with observations at only a quarterly frequency, and thus, we cannot directly detect peak-load pricing patterns.

While airlines use peak-load pricing to deal with predictable fluctuations in demand, many have argued that the airline industry is also characterized by a high degree of unpredictable variation in demand. The "stochastic demand" literature argues that when demand is uncertain, capacity is costly, and firms commit to a price ex ante, profit-maximizing behavior on the part of firms will induce a distribution of prices, rather than a single price.[4] In other words, if a firm must pay unit capacity costs regardless of whether its output is sold, then it will set the price higher in states of demand where the good is less likely to be sold. This reasoning implies that, ex ante, firms commit to distinct prices in line with their expectation over future demand states. While this literature is growing, to our knowledge, models of pricing in environments characterized by stochastic demand fluctuations have not yet produced clear, empirically testable predictions. For this reason, as well as because of the relatively low frequency of our data, we are not able to detect the effects of stochastic

[3]See Lott and Roberts (1991) and Dana (1999) for a discussion of "peak-load pricing" in the airline industry.

[4]Prescott (1975), was the first to address this issue in the economic literature. Eden (1990) formalized Prescott's example, but did so in a setting of perfect competition. Dana (1999) extended Eden's model to monopoly and oligopoly market structures.

variation in demand on ticket prices in our analysis.

2.2 The Relationship Between Price Discrimination and Competition

Numerous studies have analyzed the relationship between price discrimination and market structure in the airline industry. Using the predecessor to our data set, Borenstein (1989) and Borenstein and Rose (1994) both found evidence of decreasing price dispersion with increasing market concentration.[5] Borenstein (1989) performed a series of cross-sectional regressions of the 20th-, 50th-, and 80th-percentile prices on a measure of market concentration (Herfindahl index) and on a host of other independent variables meant to control for cost and quality factors. He found that increasing route concentration raised the 20th- and 50th-percentile prices on average, but decreased the 80th-percentile price. Borenstein and Rose (1994) took a slightly different approach in estimating the effect of market concentration on price dispersion. Using a cross-section of individual airline ticket prices, they calculated Gini coefficients for each airline-market pair. Regressing this measure on market concentration (Herfindahl index) and other factors using instrumental variables, they found a negative and statistically significant effect, indicating lower price dispersion in more-concentrated markets.[6]

These studies were motivated by theoretical work demonstrating that price discrimination need not be confined to highly concentrated markets. Borenstein (1985) provides an early model of price discrimination arising from sorting by brand preference. He argued that consumers are heterogeneous both in their willingness to pay for a product and in their loyalty to specific brands in non-monopoly markets. If firms can segment consumers based on their willingness to pay, then competition and the lack of barriers to entry will almost never prevent the occurrence of price discrimination. "In fact, when a usable sorting mechanism exists, a firm could be forced to discriminate to avoid losses when competing with other discriminating firms" (p. 381). Using a spatial model of monopolistic competition,

[5]These studies try to control for fluctuations in peak-load pricing, leaving the remaining dispersion in prices along a given route mostly attributable to price discrimination. For this reason, the terms "price dispersion" and "price discrimination" are used interchangeably throughout these analyses. Borenstein (1989) used ticket and price data from the Databank 1A (DB1A) of the Department of Transportations's Origin and Destination Survey for the third quarter of 1987. Borenstein and Rose (1994) used data from the DB1A for the second quarter of 1986.

[6]Hayes and Ross (1998) extended the study by Borenstein and Rose (1994), measuring dispersion using three different statistics: the Gini coefficient, the Atkinson index, and the entropy index. Hayes and Ross found that price dispersion is mostly due to peak-load pricing and the fare wars of the early 1990s, although some dispersion is also attributed to price discrimination practices and competition from Southwest Airlines.

Borenstein compared sorting of consumers based on reservation prices — the traditional concept of price discrimination — with sorting based on the strength of brand preferences. He found that the two different types of price discrimination lead to significantly different effects on equilibrium sales, number of firms (brands), and welfare.[7]

In response to these findings, Gale (1993) developed a theoretical model of airline price discrimination to evaluate the effects of market concentration on price discrimination. His model shows a greater difference between advance-purchase fares and unrestricted fares in an equilibrium characterized by a non-cooperative duopoly than in an equilibrium characterized by a monopoly. The increased dispersion in the non-cooperative duopoly is due to competition between the two airlines for consumers who are more time sensitive. Thus, the results of his theoretical model seem to accord well with Borenstein and Rose's empirical finding of a negative relationship between market concentration and price discrimination.

Stavins (1996) also documents a negative relationship between concentration and price discrimination. The novel aspect of her data, taken from the *Official Airline Guide*, was the fact that it included the time of ticket purchase (ranging from 35 to 2 days in advance of departure), as well as a host of ticket characteristics. Specifically, her data included four types of restrictions: advance-purchase requirements, cancellation penalties, Saturday-night stay-over requirements, and "other," unspecified restrictions.[8] Stavins's first finding was that, ticket restrictions, in the form of Saturday night stay-over and advance-purchase requirements, have a significant and negative effect on prices. Her other main finding was that, holding market share constant, higher market concentration leads to less price discrimination on a particular route. Stavins interpreted this as evidence that, as more carriers operate in a given market, competition for leisure travelers increases, while fares charged to business travelers remain essentially unchanged.[9]

2.2.1 The Monopoly Effect and the Brand Effect

The literature we discuss shows that there are two distinct effects of the degree of competition on price discrimination. We call them the "monopoly effect" and the "brand effect."

[7] Holmes (1989) also focuses on price discrimination in non-monopoly markets. Like Borenstein, Holmes distinguishes between discrimination due to heterogeneity in industry-demand elasticities (reservation prices) and discrimination due to differences in cross-price elasticities (brand preferences).

[8] By contrast, the DB1A and DB1B databases do not contain information regarding either the date of ticket purchase or ticket restrictions. However, these databases contain an enormous number of domestic airlines and markets, while Stavins's cross-sectional data encompassed only 12 different routes.

[9] The relationship between competition and price discrimination has also been studied in other industries besides airlines. Busse and Rysman (2005), for instance, find that competition acts to increase price dispersion in *Yellow Pages* advertising.

The monopoly effect is the impact that the degree of competition in a market has on a carrier's ability to maintain a markup over marginal cost. A monopolist can set a high markup without a concern for other firms' undercutting its price, while a competitive firm must price at marginal cost to avoid losing all of its customers. Hence, in a market somewhere between these two extreme cases, we expect an increase in competition to lessen the degree to which firms can set different prices between different types of consumers. The monopoly effect therefore predicts that an increase in competition acts to reduce the markup associated with tickets purchased by relatively price-inelastic business travelers, to a level more in line with the prices charged to relatively price-elastic consumers, resulting in less price dispersion and hence a lesser degree of price discrimination. The monopoly effect is equivalent to the textbook treatment of firms sustaining markups in monopoly markets.

The brand effect, however, is based on the fact that an imperfectly competitive market may also be characterized by heterogeneity in cross-price elasticities of demand. In other words, the market may contain customers who differ in their degree of brand loyalty. If consumers who purchase fares in the upper part of the price distribution, as a result of having low elasticities of demand for air travel (business travelers), have higher brand loyalty than consumers who purchase fares in the lower part of the distribution (leisure travelers), then an increase in competition may result in a larger decrease in markups in the lower portion of the price distribution than in the higher. The brand effect therefore predicts that an increase in competition will pull the lower part of the price distribution down, while the top portion remains relatively fixed, resulting in more price dispersion and a larger degree of price discrimination.

2.3 The Role of Route-Specific Consumer Heterogeneity

Borenstein (1985) stressed that airlines must develop a usable sorting mechanism in order for airlines to discriminate among customers with different brand preferences. Frequent-flyer programs (FFPs) are an example of one such mechanism that airlines have developed to induce and exploit brand loyalty.[10] An important characteristic of FFP programs is that, through a principal-agent structure, they were created to have a greater impact on business travelers than on leisure travelers. Since employers are often not willing to absorb the associated costs of monitoring flight costs, FFPs often lead employees to seek the benefits of staying with one airline to reap the FFP rewards for themselves. In our context, the fact that FFPs affect business travelers to a greater extent than leisure travelers implies that

[10] For a detailed discussion of FFP programs and their effects see Yang and Liu (2003).

price-discrimination practices that take advantage of heterogeneous brand preferences will complement price-discrimination practices that exploit heterogeneous reservation prices. Thus, one might expect to see evidence of both types of price discrimination in markets with significant numbers of business travelers. That is, if there are significant numbers of distinct types of customers on a given route, an airline will likely have ample opportunity to price discriminate between types. In contrast, if one type of consumer dominates a route, implying little heterogeneity in elasticities or preferences over brands, then airlines will be unable to use price-discrimination strategies effectively.

Previous empirical studies of airline prices have attempted to control in various ways for the mixture of business travelers and tourist travelers in a particular market. Borenstein (1989) constructed a tourism index at the SMSA level, using the ratio of hotel income from tourist customers to total personal income, and included the index as a control variable in his estimates of a linear pricing equation.[11] Brueckner, Dyer, and Spiller (1992) and Stavins (1996) included the absolute difference in mean January temperatures between origin and destination as a proxy for tourism in their reduced-form pricing regressions.

In all of these studies, measures of tourism are allowed to affect only the level of prices or price dispersion. However, if price discrimination is a main determinant of price dispersion, then the extent of consumer heterogeneity on a route should also affect the manner in which competition affects the level of prices. The analysis we perform shows that the dynamics of price distributions in relation to competition appear to vary more across markets that have different compositions of business versus leisure travelers than across markets with more homogeneous travelers. In Section 3, below, we discuss our methods for distinguishing between routes with a significant proportion of business travelers and routes dominated mostly by leisure travelers.

[11] Borenstein and Rose (1994) use the same index as a control variable in their estimates of price-dispersion regressions.

3 Data

3.1 Data Sources and Variable Construction

Our study focuses on domestic, direct, coach-class tickets from nine major carriers over the period 1993 to 2006.[12] "Low-cost carriers"[13] (LCCs) and regional carriers are excluded to maintain consistency with the samples used by Borenstein and Rose, as well as to adhere to the practice of many airline economists, who view these types of airlines as inherently different in both cost structure and pricing strategies from the traditional "hub-and-spoke" carriers.[14] However, our analysis accounts for the effects of competition from LCCs and regional carriers on the carriers in our sample. Ticket prices are obtained from the DB1B database, which is a 10-percent sample of all domestic tickets issued by airlines.[15] In addition to ticket prices, the DB1B includes other quarterly itinerary information, such as origin and destination airports, passenger quantities, number of stops (plane changes), and fare class.[16] Any tickets believed to be frequent-flyer tickets are eliminated.[17]

We construct a panel, where an observation is a flight conducted by a specific airline, between an origin and destination airport (route), in a specific time period (year and quarter). For example, a United Airlines coach-class ticket, non-stop from Philadelphia (PHL) to Chicago (ORD) in the first quarter of 1999 is considered an observation in our data. Our non-stop ticket data include both one-way flights and round-trip flights. The DB1B contains numerous itineraries and fares for the same flight by the same carrier, reflecting the quarterly frequency of the data, as well as the many different fares found within the same

[12]Specifically, we merge three sources of data in this study, all from TranStats, the Bureau of Transportation's (BTS) online collection of databases. The data on ticket prices come from the Airline Origin and Destination Survey (DB1B). Some of the market/route characteristics that are used in our empirical regression models originate from the T-100 Domestic Segment database, which is a component of Form 41. The carriers in our sample include United, US Airways, Delta, American, Alaskan, TWA, Continental, Northwest, and America West.

[13]The list of LCCs, obtained from Ito and Lee (2003a), includes Air South, Access Air, AirTran, American Trans Air, Eastwind, Frontier, JetBlue, Kiwi, Morris Air, National, Pro Air, Reno, Southwest, Spirit, Sun Country, ValuJet, Vanguard, and Western Pacific.

[14]For a more detailed discussion of LCCs, see Ito and Lee (2003a) and Goolsbee and Syverson (2005).

[15]The DB1B is the online version of the original DB1A database, which was used in many of the studies referenced above. Unfortunately, the DB1B only has data going back to 1993, which prohibits us from using the same data as these previous studies.

[16]There are three different sub-components to the DB1B data set. They are market data, coupon data, and ticket data; and we combine variables from all three. For further reference, see the BTS's website http://www.transtats.bts.gov.

[17]Tickets obtained using frequent-flyer miles are typically charged handling fees between \$5 and \$15, depending on the airline. Thus, to be safe, we eliminate any tickets with prices under \$20. These criteria are commonly used in the literature in an effort to control for ticket quality.

fare class, on the same flight, at a given point in time. Thus, our data comprise distributions of prices for carrier-route itineraries. Two typical price distributions are displayed in Figure 1.[18]

We obtain additional route characteristics to supplement the DB1B from the Bureau of Transportation Statistics'(BTS) T-100 Domestic Segment Database, which is derived from Form 41 (Traffic). This database contains domestic, non-stop segment data reported by all U.S. carriers, including passengers transported, origin, destination, aircraft type, available capacity, scheduled departures, departures performed, and aircraft hours. Since these are segment data, they are compatible with the data on direct flights that we use from the DB1B. One significant difference between the two data sets is with respect to passenger counts. The T-100 data set includes observations on enplaned passengers, encompassing passengers originating and ending their trips at the origin and destination airports as well as passengers who are connecting to and from other flights at the airports. The DB1B data on direct flights, however, include only passengers who are originating and ending their flights at the respective origin and destination airports.[19] Appendix A contains a comprehensive discussion of data sources and variable construction.

3.2 Summary Statistics of Price Dispersion and Competition

We define the ticket price as a single-direction fare, so that the prices of one-way flights are exactly as listed in the itineraries, while the prices of round-trip flights are one-half of the prices listed in the itineraries.[20] Table 1 displays summary characteristics of the data after merging the DB1B data on direct flights with the T-100 Domestic Segment data. The first panel of Table 1 provides the summary statistics for the first quarter of our sample (1993:Q1) and the second panel displays the statistics for the last quarter of our sample (2006:Q3).

Our analysis follows Borenstein and Rose (1994) and other studies of airline pricing in

[18]Each quarter of the DB1B database contains a very large amount of data. For example, 1993:Q1 contains approximately 4.8 million coupons, where a coupon essentially identifies a segment of travel (that is, a one-way flight from Boston to Las Vegas that stops in Chicago would have two coupons, BOS-ORD and ORD-LAS). We hypothesize that, as a result of restrictions on computing power, the studies mentioned above that used the DB1A were forced to work with single cross-sections of the data. Thanks to increases in computing power, we were able to construct and work with 55 quarters of data spanning more than 13 years.

[19]Another difference is that the T-100 data set includes information on almost every domestic flight segment flown by domestic carriers and thus contains more routes than the DB1B data.

[20]This is identical to the methods of Borenstein and Rose (1994).

focusing on the Gini coefficient.[21] As Borenstein and Rose discuss, the expected absolute difference between two ticket prices drawn randomly from the population is equal to the Gini coefficient. For example, a Gini coefficient of 0.25 on a given route and carrier implies an expected absolute price difference of 50 percent of the mean fare on that route.

Table 1 suggests a substantial amount of dispersion in prices in our sample.[22] The average Gini coefficient is 0.22 in both 1993 and 2006, which corresponds to an expected fare difference of 44 percent of the mean fare for two randomly selected passengers on a given carrier and route. This is larger than the 36-percent difference obtained by Borenstein and Rose (1994) using data from 1986, suggesting that price dispersion has increased as the airline industry has evolved into its present state.

The second and third row of each panel in Table 1 contain information regarding levels of competition. The average number of competitors per route has decreased between 1993 and 2006 from approximately 3.0 to 2.6.[23] We also calculate an additional statistic, which can be interpreted as the "effective" number of competitors per route. It is simply the inverse of the Herfindahl index calculated using passenger shares, and therefore takes into account the degree of firm concentration on a given route.[24] This statistic displays no discernible change. For all flight types, the average effective number of competitors remained virtually constant between the first quarter of 1993 and the third quarter of 2006.

3.3 "Big-City Routes" vs. "Leisure Routes"

The theory of price discrimination suggests that price dispersion depends on the mixture of customer types on a given route. In particular, if price discrimination is a main determinant of price dispersion, then routes with multiple types of customers should be more affected by competition than routes with homogeneous customer bases. In our analysis we therefore compare the effects of competition on price dispersion, taking into account the observable types of travelers purchasing tickets. We presume that tickets for routes to leisure

[21] See, for example, Borenstein and Rose (1994). Hayes and Ross (1998) use the Gini coefficient and Atkinson's Index, as well as Theil's index. For a comprehensive discussion of the advantages and disadvantages of different dispersion statistics we refer the reader to Cowell (1995). We provide summary statistics

[22] See Appendix C for an expanded table listing other measures of price dispersion, along with other variables of interest.

[23] The number of competitors is taken from the DB1B and is the total number of carriers—legacy-carriers, low-cost carriers, and regional carriers—operating on a given route in a given quarter.

[24] This statistic is calculated using passenger quantity information from the DB1B. We also calculated market shares and a Herfindahl index based on passenger shares (enplaned) as well as flight shares from the T-100 Segment Data, with all three measures sharing a high correlation. If plane sizes and load factors do not differ substantially across carriers on a given route, we would expect the two types of Herfindahl indexes to be similar.

destinations are mainly bought by passengers with high elasticities of demand and hence low reservation prices—leisure travelers. We also presume that tickets for routes between big cities are purchased by leisure travelers as well as by passengers with high reservation prices — business travelers. Specifically, we segment our data into "leisure routes"—routes with mainly price-sensitive leisure travelers—and "big-city routes"—routes with both leisure and price-insensitive, business travelers.

We take the following steps in distinguishing between big-city and leisure routes: For each airport in our data, we calculate the ratio of accommodation earnings to total non-farm earnings corresponding to the Metropolitan Area (MA) containing that particular airport for each year over the period 2001–2004, then take the median value.[25] We then sort these ratios in descending order and label all routes that include an airport in an MA above the 85th percentile as a leisure route.[26] Table 2 displays the airports in MAs that have accommodation-to-nonfarm earnings in the 85th percentile or above. In addition to the airports in the 85th percentile, we include a few airports from U.S. territories for which we have no MA earnings data, such as San Juan, St. Croix, and St. Thomas.[27]

Our criterion for choosing the big-city route sample is even simpler than our criterion for choosing the leisure sample. We classify a route as "big city" if that route contains both an origin and a destination airport located within the 30 largest MAs in the United States (in terms of population).[28] These MAs along with their respective airports and populations are shown in Table 3.[29] Our assumption is that there is a large proportion of business travelers on routes between large cities.

Figure 1 displays the distribution of coach-class fares of a leisure route—US Airways route from Philadephia (PHL) to Orlando (MCO)—as well as a big-city route—United Airlines route from Philadelphia (PHL) to Chicago (ORD)—for 1999:Q1. The big-city route shows more price dispersion than the leisure route, with Gini coefficients of .303 and .248, respectively. This figure also shows that these two routes are characterized by different price

[25]We obtained these data from the Bureau of Economic Analysis (Table SA05 - Annual Personal Income by Major Source and Earnings by industry). Since the BEA changed its industry classification system in 2000, and the new codes (NAICS) are not readily compatible with the old (SIC), we were unable to calculate this ratio for the entire span of our data.

[26]Memphis is the lone exception. We do not have a good explanation for Memphis's large ratio of accommodation earnings to nonfarm earnings (over 2 percent), and thus believe that it could be a data coding error.

[27]The BTS's definition of domestic includes all U.S. territories.

[28]There are a few exceptions to this criterion, and they are identified in the last column of Table 3. We did not include airports located in Miami, Fort Lauderdale, San Diego, Tampa Bay, or Orlando on the basis that all of these areas are characterized by significant tourism industries as well as by high amounts of business activity.

[29]Population figures are taken from the Census Bureau, and correspond to July 1, 2005.

distributions. The leisure route appears to have a unimodal distribution of prices, indicating that a majority of the tickets were sold for around 100 dollars. We presume the carrier was targeting one type of consumer—leisure travelers. The big-city route, however, appears to have a bimodal price distribution, indicating that the carrier sold a large portion of tickets for around 100 dollars and also a large portion for around 450 dollars. We believe that the carrier here was targeting its prices at both leisure and business travelers.

In columns (2) and (3) of Table 1 we display summary statistics for our samples of big-city and leisure routes for the first and last quarter of our sample. The table shows that in both quarters the mean Gini coefficient is larger in the big-city route sample. In order to show a broader range of the Gini coefficients in our data set, we plot density functions in Figure 2 corresponding to the Gini coefficients calculated from our route-specific price distributions. To construct this figure, we calculate the Gini coefficient at each date for each route, and then segment these variables into various samples of potential interest. The first panel (upper-left) displays two density functions. One is the density of the Gini coefficients of all monopoly routes in the sample, and one is the density of all other routes, which we label competitive routes.[30] The density plot shows that monopoly routes are characterized by more price dispersion (0.27 median Gini coefficient) than competitive routes (0.24 median Gini coefficient). The second panel (upper-right) displays similar density plots after segmenting the sample between big-city and leisure routes. Here we see a much larger difference in price dispersion, as routes with a large proportion of leisure travelers are characterized by much less price dispersion than routes with larger proportions of business travelers, with a median Gini coefficient of 0.20 versus a median Gini coefficient of 0.28, respectively. Finally, the bottom panels in Figure 2 partition both big-city and leisure routes into competitive and monopoly routes. Leisure routes, which we believe are not patronized by many price-inelastic business consumers, show a negligible difference in dispersion between competitive and monopoly routes. However, when we restrict the sample to big-city routes, we again see a significant difference in dispersion between monopoly and competitive routes.

These plots suggest that the composition of consumer types on a given route may influence the effect of competition on price dispersion. That is, they show that competition may have a larger effect on price dispersion when there are distinct types of consumers purchasing tickets. This in turn suggests not only that price dispersion is at least partly attributable to price discrimination, but also that competition may reduce the ability of a carrier to price discriminate.

[30] In this figure, we define monopoly routes as routes on which one firm's average market share for each quarter over the entire sample period is greater than 0.95.

3.4 Example Routes

In this section we present a few graphical examples of the dynamic price distributions that are representative of many of the leisure and big-city routes in our respective samples, paying particular attention to the role of competitive forces. The bottom panel of Figure 3 shows the 90th and 10th percentiles of ticket prices for three carriers—United Airlines, American Airlines, and US Airways—operating on non-stop flights from Philadelphia (PHL) to Chicago (ORD) over the entire sample period. This route is between two large metropolitan areas, and thus we expect it to be characterized by large numbers of both business and leisure travelers. In the figure, the 10th- and 90th-percentile prices of all three carriers appear to move in tandem with one another, suggesting that these three carriers are either explicitly colluding or are paying close attention to one another when setting quantities and prices. The top panel of Figure 3 shows similar plots for a leisure route in our sample—Los Angeles (LAX) to Honolulu (HNL). On this route, we do not see the same degree of correlation among the percentiles of ticket prices for the three carriers. The carriers here appear to be competing with each other in a different manner than those in the previous example.

The type of route also appears to play an important role in how the entry and exit of carriers, especially LCCs and regional carriers, affect the price distributions of the carriers in our sample. The bottom panel of Figure 4 displays all of the price deciles from the Philadelphia (PHL)—Chicago (ORD) United Airlines route. Also plotted in this figure are the 90th percentiles of two LCCs, ATA and Southwest, and a regional airline, Midway Airlines.[31] Entry and exit by these carriers appear to have had significant impacts on the dynamics of the price distribution. Entry by Midway Airlines and Southwest Airlines appears to have pulled down the upper percentiles of the distribution to a much greater extent than the lower percentiles, thereby reducing price dispersion.[32] The top panel of Figure 4 shows the entry and exit of carriers on the leisure route operated by US Airways from Philadelphia (PHL) to Orlando (MCO). On this route, which we believe to be characterized by a more homogeneous customer base of price-elastic travelers, entry and exit seem to have had a smaller effect on price dispersion.

The examples discussed in this section are indicative of many of the routes in our big-city and leisure samples. They suggest that the type of route in question may play an important role in the relationship between pricing and competition. As Figure 3 suggests,

[31] Midway and Southwest entered these routes through Midway Airport (MDW), not O'Hare International Airport (ORD).

[32] For a discussion of the threat of entry by Southwest Airlines into incumbent carriers' routes, see Goolsbee and Syverson (2005).

depending on the route, incumbents may react differently to one another in setting prices and quantities. Furthermore, when price discrimination is a viable option for carriers, as on the big-city routes, competition seems to have a large effect on price dispersion.

4 Cross-Sectional Analysis

The pricing study of Borenstein and Rose (1994) explored the effect of market concentration on price dispersion using a cross-section of data for 1986. While our sample period is 1993-2006, and although we lack some of the cost variables used by Borenstein and Rose, we find that the cross-sectional regressions for the vast majority of the 55 quarters in our sample lead to very similar findings.[33] Specifically, we find a negative and significant relationship between route concentration and price dispersion. Furthermore, we find evidence of a positive and statistically significant relationship between the number of carriers operating in a market and the extent of price dispersion in that market. Thus, our cross-sectional analysis updates and extends the findings of Borenstein and Rose.

In this section we discuss two measures that we use as a proxy for competition in the cross-sectional analysis: route concentration and the number of carriers. We subsequently perform IV regression analysis.

4.1 Measuring Competition in the Cross Section

In exploring the effect of competition on price dispersion, we have two concerns. First, we want to use measures that are consistent with those in prior studies, so that we can replicate previous empirical analysis on our data set. Second, we want to have a battery of measures of competition so that we can examine whether our empirical findings are robust. In the cross-section analysis, we use two different methods to control for the level of competition on a given route.

4.1.1 Route concentration

Our first method corresponds exactly to the method used by Borenstein (1989) and Borenstein and Rose (1994), in which we measure overall route competition with the log of market concentration of the route (the Herfindahl Index), $\ln HERF_j$, while controlling

[33] Borenstein and Rose obtained supplemental data from the *Official Airline Guide* to construct some of their cost variables, while we were unable to obtain access to this source. However, we were able to obtain some of those variables from the T100-Segment database, which we will discuss in more detail below.

for the carrier's specific market share on the route, $\ln MKTSHARE_{ij}$.[34] Here, the index i corresponds to the carrier, and j corresponds to the route. In this method, market share is isolated from route concentration in order to hold fixed the market power specific to the carrier operating on the route. That is, this method theoretically allows the econometrician to assess the effects of route concentration between routes assuming that each carrier has the same amount of market power.[35]

The expected effect from route concentration will depend on whether the monopoly effect or the brand effect dominates. If route concentration has a greater impact on expensive tickets than on cheap tickets, then the resulting increase in price dispersion will be evidence of the monopoly effect. However, if the higher percentiles are less affected by route concentration than the lower percentiles, then we expect to see a negative effect of $\ln HERF_j$. The resulting decrease in price dispersion would be attributable to the brand effect.

Since the Gini coefficient is bounded between 0 and 1, we measure price dispersion using the Gini log-odds ratio given by $G_{ij}^{lodd} = \frac{ln(G_{ij})}{1-ln(G_{ij})}$, which produces an unbounded statistic.[36] The first model is

$$\begin{aligned} G_{ij}^{lodd} &= \beta_0 + \beta_1 \ln MK\widehat{TSH}ARE_{ij} + \beta_2 \ln \widehat{HERF}_j + \beta_3 \ln F\widehat{LTTO}T_j \\ &+ \beta_4 \ln TOURIST_j + \beta_5 HUB_{ij} + \beta_6 SMALL_j + \alpha_i + \gamma_j + \eta_{ij}. \end{aligned} \quad (1)$$

Like Borenstein and Rose, we specify that carrier effects, α_i, are fixed, and that route effects, γ_j, are random. The variable $TOURIST$ is the maximum of the ratio of accommodation earnings to total non-farm earnings for the origin and destination airports; HUB is a dummy variable indicating whether the origin or the destination is a hub airport for the given carrier; $SMALL$ is a dummy variable equal to one if the route does not include a big city; and $FLTTOT$ is the total number of flights on a given route, and is included as a proxy for market density.[37] In Appendix D we provide a table that compares the variables

[34] The proportion of total passengers (in a given quarter) originating on route j on carrier i, is used as a proxy for market share, $MKTSHARE_{ij}$, which is subsequently used in the calculation of the Herfindahl index. Borenstein and Rose (1994) use flight shares—the proportion of total flights on route j on carrier i—as the relevant market share variable. Passengers shares and flight shares are highly correlated; however, passenger shares show more variation over time on a given route. We therefore thought it was the more relevant variable for the panel analysis. No results in this paper change using flight shares.

[35] Evans and Kessides (1993) find that airfare is not correlated with market share; however, Borenstein (1989) finds a statistically significant effect of market share on price.

[36] No results change using the log of the Gini coefficient. See Hayes and Ross (1998) for further discussion.

[37] Borenstein and Rose (1994) also include variables that control for weekly variation in fleet utilization rates and airport-capacity utilization rates in order to control for predictable or "systematic" peak-load

used in our model with those used in the analysis of Borenstein and Rose (1994).

4.1.2 Number of carriers

As a more general control, our second method follows Berry (1992), where we simply include the log of the total number of carriers operating along the route, $\ln N_j$, as a way to capture any variation in the markup due to variation in the general level of competition. If the monopoly effect dominates the brand effect, an increase in the number of competitors on a given route will decrease the higher-percentile prices more than the lower-percentile prices, decreasing the overall degree of price dispersion on that route. On the other hand, if the brand effect dominates, an increase in the number of competitors will decrease prices in the lower part of the distribution by more than those in the upper part, thereby increasing the overall degree of price dispersion.

The second method, which proxies for competition with the number of competitors operating on the route, is

$$\begin{aligned} G_{ij}^{lodd} &= \beta_0 + \beta_1 \ln \widehat{N}_j + \beta_2 \ln \widehat{FLTTOT}_j \\ &+ \beta_3 \ln TOURIST_j + \beta_4 HUB_{ij} + \beta_5 SMALL_j + \alpha_i + \gamma_j + \eta_{ij}. \end{aligned} \tag{2}$$

We instrument for N, $MKTSHARE$, $HERF$, and $FLTTOT$, using the same instruments as used in Borenstein and Rose; these include distance, population, total passengers, and two instruments constructed by the authors.[38]

4.2 Estimation Results

We perform cross-sectional regressions on each of the 55 quarters in our sample in order to directly compare our results with those of Borenstein and Rose. For the sake of brevity, we report the first quarter estimates in six-year spans between 1993 and 2005 in Table 4. Table 4 reports results from both models for the three time periods specified above.

Consistent with the results found by Borenstein and Rose, the results obtained in Table 4 suggest that more competitive routes are characterized by more price dispersion. This result

pricing, using data from the *The Official Airline Guide* that are unavailable to us. We were able to construct seat-capacity utilization rates at a monthly frequency, using data from the T100-Segment database, but this variable's effect is not significantly different from zero in any of the specifications we use, nor does it affect the other estimates in the model; thus, it is omitted from the estimations reported below.

[38]Please refer to Appendix B for a detailed definition of all of the variables as well as the instruments.

can be seen in the negative and statistically significant coefficient estimates on $\ln\widehat{HERF}$ and also in the statistically significant positive coefficient estimates associated with $\ln\widehat{N}$, found in 1993:Q1 and 1999:Q1. These results are consistent with most of the quarters in our sample. In 44 of 55 quarters we obtain negative estimates on $\ln \widehat{HERF}$, and in 40 quarters we obtain positive estimates on $\ln \widehat{N}$, the majority of which are statistically significant at the 10-percent level or lower.

Borenstein and Rose interpret this positive relationship between price dispersion and competition as evidence of the brand effect. One difference between our results in Table 4 and Borenstein and Rose's results is that we find a statistically significant, positive effect of a carrier's market share on price dispersion in all three of the time periods. In the first quarter of 1993, a 10-percent increase in market share is estimated to increase price dispersion by 1.6 percent on average, *ceteris paribus*. This finding is consistent with the monopoly effect, as an increase in a firm's market power is expected to increase its ability to segment the market. In contrast, Borenstein and Rose find a positive, but not statistically significant effect from market share on price dispersion in their estimation.[39] Overall, our updating and extension of Borenstein and Rose (1994) is consistent with the results of their analysis..

5 Panel Analysis

In this section, we exploit the panel dimension of our data in order to assess the effects of competition on price dispersion while controlling for time-invariant, route-specific factors. We use a fixed-effects approach, which exploits only the time-series variation along a specific route in the estimation routine. Hence, we estimate the effect of competition on price dispersion while looking at changes in the competitive structure of a given route *over time* rather than differences in competitive structures *across routes*.

We use two different approaches to analyze the effects of competition on the distribution of prices charged by airlines over the course of our sample. First, as in the cross-sectional analysis, we use the Gini coefficient as the dependent variable to proxy for price dispersion. Second, we estimate a series of reduced-form pricing equations, similar in spirit to Borenstein (1989), in which we use the 90th and 10th percentiles of the price distribution as our dependent variables.[40]

[39] One reason why our standard error on the coefficient on market share is smaller than that obtained by Borenstein and Rose might be that we use a larger sample for each cross section.

[40] In order to determine initially whether time-invariant, route-specific effects would be important in our context, we performed Hausman tests for each sample and specification of our model. In all cases the null hypothesis of zero correlation between the residuals and the vector of explanatory variables was soundly

5.1 Measuring Competition in the Panel

We measure competition in the panel using two methods similar to the cross-section, and introduce a third measure that accounts for the type of other carriers operating on the route. One major difference between the cross-sectional estimation and the panel estimation is the way in which we control for carrier-route-specific factors. In the cross-section, we used a range of carrier-route- and route-specific covariates. In the panel, however, we include carrier-route dummies which control for *all* of the characteristics specific to the carrier-route observation. Therefore, many of the carrier-route- and route-specific variables included in the cross sectional estimation do not belong in the fixed-effects panel regressions. In particular, it is not necessary to control for the market power of the carrier since the effects of competition are no longer being compared between routes. Any variation in the market share of the carrier over time on a given route is generally due to a change in the degree of competition on the route, which is controlled for with route concentration. Similarly, we no longer need to keep fixed the total number of flights on the route, since most variation in this variable over time is due to variation in the number of competitors.[41]

5.1.1 Route concentration

Our first model measures the degree of competition with the Herfindahl index of the route:

$$G_{ijt}^{lodd} = \theta_0 + \beta * \ln\widehat{HERF}_{jt}\alpha * X_{ijt} + \gamma_t + \upsilon_{ij} + \nu_{ijt}. \tag{3}$$

5.1.2 Number of carriers

Our second model proxies for competition with the number of carriers:

$$G_{ijt}^{lodd} = \theta_0 + \beta * \ln\widehat{N}_{jt} + \alpha * X_{ijt} + \gamma_t + \upsilon_{ij} + \nu_{ijt}. \tag{4}$$

rejected. This suggests that controlling for time-invariant, route-specific effects is very important and validates our choice of fixed-effects estimation.

[41] Indeed, there was very strong correlation between $MKTSHARE$, $HERF$, and $FLTTOT$ on most of the routes. As a robustness check, we performed the panel regressions with $\ln MKTSHARE$ included as a right-hand-side variable. Its coefficient was positive and significant in all regressions, and the estimated coefficient on $\ln HERF$ was subsequently lowered, but never to a negative and significant value.

5.1.3 Types of carriers

Our third method controls for the effects of competition between different types of carriers operating in a given market. Over the past 10 to 15 years, low-cost carriers have emerged as an important source of competition for the traditional legacy carriers. Thus, we have reason to believe that low-cost carriers may explain a significant amount of the variation over time in the competitive structure of markets during this period. For this method, we broaden the definition of route to a "city pair," which groups airports in a given metropolitan area together, and defines a route as travel between two cities. Many LCCs entered markets through different airports (within the same metropolitan area) than those used by the legacy carriers. In order to capture the presence of these LCCs in a given market, we need to define routes as flights between cities, not between airports. Thus, we do not treat routes involving airports located in the same city as separate from one another.[42] For each of the types of carrier—legacy carriers, low-cost carriers (LCCs), and regional carriers—we include dummy variables, $LEGACYdum_{jt}$, $LCCdum_{jt}$, $REGdum_{jt}$, indicating whether or not an airline of that type is present in the city pair. As Southwest Airlines is the largest and most prevalent LCC in our sample, we isolate Southwest from the other LCCs, with $SWdum_{jt}$. Regional carriers include all other non-legacy carriers. As in Goolsbee and Syverson (2005), we also account for any variation in the markup due to the threat of entry by an LCC. Specifically, we include dummy variables indicating whether an LCC has a presence at both origin and destination cities, but is not currently operating on the route. We again split the LCCs between Southwest, $SWthreat_{jt}$, and other LCCs, $LCCthreat_{jt}$. The resulting model is

$$\begin{aligned} G^{lodd}_{ijt} &= \theta_0 + \beta_1 * LCCdum_{jt} + \beta_2 * SWdum_{jt} + \beta_3 * LEGACYdum_{jt} \\ &+ \beta_4 * REGIONALdum_{jt} + \beta_5 * LCCthreat_{jt} + \beta_6 * SWthreat_{jt} \\ &+ \alpha * X_{ijt} + \gamma_t + \upsilon_{ij} + \nu_{ijt}. \end{aligned} \tag{5}$$

5.2 Panel Estimation Description

In all models, i indexes the carrier, j the route, and t the time period.[43] We control for carrier-route fixed effects with υ_{ij}, and our controls, X_{ijt}, include quarter dummies and

[42] For example, a flight from O'Hare (ORD) to Philadelphia is considered to be on the same route as a flight between Midway (MDW) and Philadelphia.

[43] We do not include HUB, $TOURIST$, or $SMALL$ in these models, as these variables are time-invariant.

dummy variables indicating whether the airline is bankrupt in period t. We control for important exogenous cost and demand effects through a full set of time dummies, γ_t.[44] In the first two models, we include the same instruments as in the cross-sectional estimation; however, it is necessary to remove time-invariant variables such as distance.[45] In addition, we cluster our standard errors by route, as in Goolsbee and Syverson (2005), in order to control for both serial correlation and correlation between the pricing decisions of multiple carriers on the same route, which is apparent in Figure 3.[46] There is, of course, a possibility that the residuals are correlated across different routes within the same airline. For instance, an unaccounted-for, carrier-specific shock could affect all prices on an airline's network, producing incorrect standard errors associated with our estimates. Although we cannot account for all carrier-specific shocks, we believe that including bankruptcy conditions as a control is likely to capture some of the important carrier-specific shocks.[47]

We also estimate the same models, substituting the Gini coefficient with the 90th- and 10th-percentile prices of carrier i, route j, in period t. While using the Gini coefficient seems to be the more popular approach in the literature, we see both advantages and disadvantages to this method compared with using certain percentiles of the price distribution as the left-hand-side variable. On the one hand, the use of a dispersion statistic as a dependent variable allows for a more direct interpretation of the effects from the explanatory variables on price discrimination and dispersion. The disadvantage of this approach comes from the restrictive nature of using a single statistic to summarize an entire distribution. For example, the Gini coefficient places more emphasis on the middle part of the distribution, and is not as sensitive to the tails of the distribution. Also, analyzing the percentiles sheds more information on the change in the shape of the price distribution. For example, price dispersion can increase because of a rise in the upper portion of the price distribution relative to the lower portion, or it can increase because the lower portion falls by more than the upper portion. For this reason, we believe that studying a range of percentiles of the price distribution may be more informative than concentrating on only a single statistic. However, using percentiles of the price distribution as dependent variables does not allow for a direct and formal analysis

[44]We also included average variable cost (obtained from the BTS's P-52 database) of the carrier as a control, but it did not affect our estimates. We chose to leave it out because of endogeneity concerns.

[45]Due to the lack of a sufficient number instruments, we do not instrument for the competition dummy variables in the "types of carriers" model. However, since high markups will induce more competitors to enter, any endogeneity bias will be positive.

[46]Strategic behavior on the part of airlines has been well documented in the literature. For a few examples, see Alam, Ross, and Sickles (2001), Berry (1990), and Brueckner and Spiller (1991).

[47]See Borenstein and Rose (1995) for an analysis of the effects of bankruptcy on pricing in the airline industry.

of price dispersion. We therefore perform the same sets of estimations on both types of dependent variables in order to obtain a clearer picture of the effects of market structure on prices.

5.3 Estimation Results

Table 5 contains estimation results of all three models using the Gini coefficient as the dependent variable, while Table 6 shows results using the 90th and 10th percentiles of the price distribution as the dependent variables for the big-city sample and the leisure sample.[48] We report results for all direct routes operated by legacy carriers in our 13-year sample. This sample includes 82,855 observations, covering 4,728 carrier-route observations, and 2,752 distinct routes. In Table 5, we include results for all flights in our sample (panel 1), as well as results for our big-city route sample (panel 2) and our leisure route sample (panel 3).[49] In Section 3 we saw that the degree of consumer heterogeneity seems to play a role in the distribution of Gini coefficients in our sample. Performing separate estimations for big-city routes and leisure routes should better reveal how the effects of competition affect carriers' ability to price discriminate. If competition does erode the ability of carriers to price discriminate, then we would expect to observe larger effects from competition on big-city routes, where we have both price-elastic and price-inelastic consumers, than on leisure routes, where we believe there to be a more homogeneous group of price-elastic consumers.

For the sample of all routes and for the big-city route sample, the effect of market concentration, $\ln \widehat{HERF}$, on price dispersion is *positive* and significant at the 5-percent and 1-percent significance levels, respectively. This effect is exactly the opposite to what we found in the cross section. In the sample of leisure routes, the effect of route concentration on price dispersion is insignificant. A look at the estimates corresponding to $\ln\widehat{HERF}$ from the percentile regressions sheds more light on this issue. For the sample of big-city routes, an increase in route concentration increases the 90th-percentile price level by over twice the amount it raises the 10th-percentile price. On the sample of leisure routes, the 90th- and 10th-percentile price levels rise by similar amounts. Thus, an increase in route concentration has a larger positive effect on the top portion of the price distribution when there are a significant number of price-inelastic business consumers in the consumer pool. When the

[48]We omit the coefficient estimates corresponding to our set of controls, X_{ijt}, for the sake of brevity. There are no unexpected findings, but the results are available upon request from the authors.

[49]The big-city route sample consists of 27,030 observations, covering 1,344 distinct carrier-route observations and 598 different routes, while the leisure flight sample consists of 15,666 observations, covering 932 distinct carrier-route observations and 529 different routes.

consumer base is more homogeneous, as in the leisure route sample, route concentration affects the top and bottom portions of the price distribution almost equally.

The results from the second model show that an increase in the number of competitors, $\ln \widehat{N}$, on a route is estimated to have a statistically significant negative impact on price dispersion in the sample of all flights. This suggests that increases in the number of competitors, on average, leads to decreases in price dispersion, which, again, is exactly opposite to the effect found in the cross-sectional regressions. The effect of $\ln \widehat{N}$ on price dispersion is significant, negative, and larger (in absolute value) for our sample of big-city routes, and smaller and not significantly different from zero in the sample of leisure routes. Focusing on Table 6, we see that the 90th-percentile price level falls by around twice as much as the 10th-percentile price level following an increase in competition on big-city routes. In contrast, on leisure routes, both percentiles fall by roughly the same proportion. Thus, similar to the results found using the first model, these results suggest that, on routes with significant heterogeneity in consumer types, an increase in competition erodes the ability of carriers to charge a high markup to price-inelastic customers. Competition thus limits the ability of the carrier to price discriminate between consumer types, which lends support to the monopoly effect.

The results from the third model also provide evidence that an increase in competition results in less price discrimination. Interestingly, the presence of an LCC or a regional carrier on a city pair has a much larger effect on a given carrier's price dispersion than does the presence of another legacy carrier. Estimates of the effects from the dichotomous variables corresponding to the presence of Southwest, $SWdum$, and other LCCs, $LCCdum$, on the city pair are negative and largely significant for the sample of all routes and the sample of big-city routes. The presence of an LCC in a market is estimated to decrease a legacy carrier's price dispersion by approximately 2 to 5 percent, depending on whether we use the full sample of routes or only the big-city routes. Regional carriers, measured through $REGIONALdum$, have a significant, negative effect on the sample of big-city routes, but not on the full sample or leisure-route sample. However, estimates of the effects from the presence of another legacy carrier, $LEGACYdum$, are all insignificant. The big-city results for the price percentiles (Table 6) verify that in each case the 90th-percentile price-level falls by more than the 10th-percentile price level, indicating that the presence of an LCC on a route severely reduces the ability of the incumbent carrier to charge a high markup to price-inelastic consumers relative to the markup they can charge to price-elastic consumers. For the sample of leisure routes, the estimates of all competition variables are insignificant, with the exception of Southwest. Estimation results for the effects of the threat-of-entry variables, $SWthreat$ and $LCCthreat$, show mixed results. The coefficient on $SWthreat$ is negative but insignificant on all samples, while the coefficient on $LCCthreat$ is positive and significant for the sample of all routes and for the sample of big-city routes in Table 6. From the percentile regression results, we can see that the estimate for $LCCthreat$ is due to a greater decline in the 10th-percentile price compared with the decline in the 90th-percentile price. Thus, our results do show some evidence of the brand effect, but only in cases where there is a threat of entry. In cases of actual entry by either Southwest or by another LCC, price dispersion falls.[50]

6 The Role of Route-Specific Factors

Our fixed-effects panel analysis shows that after controlling for time-invariant, route-specific effects, the relationship between competition and price dispersion changes sign. The coefficient estimate for $\ln\widehat{HERF}$ becomes positive, while the estimate for $\ln\widehat{N}$ becomes negative. This result suggests that time-invariant, route-specific effects may be responsible for biasing the estimates on our competition variables. In this section, we discuss a particular route-specific effect—the distance of the route—that we believe may play a significant role in creating bias in the cross-sectional estimates of the effect of competition on price dispersion.

In the cross-sectional regressions performed by Borenstein and Rose, route effects, γ_j, are treated as random and distance is included in the instrument set, since it should be a good instrument for the total number of flights on route, $FLTTOT_j$. Including distance as an instrument is valid as long as distance is not correlated with either of the error terms, η_{ij} or γ_j. In the following discussion we explain why such correlation is in fact present, and show that distance is one of the route-specific effects responsible for biasing the estimates on the competition variables in the cross section.

Letting X_{ij} denote our set of explanatory variables (which includes both exogenous and potentially endogenous variables), we can rewrite equation (1) more compactly as

$$G_{ij}^{lodd} = \beta X_{ij} + \alpha_i + \gamma_j + \eta_{ij}. \tag{6}$$

[50]As a check on the robustness of the results discussed above, we also performed the same regressions, substituting the Atkinson index for the Gini coefficient as our statistic for summarizing price dispersion (that is, the left-hand-side variable). The Atkinson index is a statistic that is often used in the inequality literature. The qualitative results from these regressions were all consistent with the results that we obtained from the Gini-coefficient regressions. For brevity, we do not include these results, but they are available from the authors upon request.

Letting Z_{ij} denote the instrument set for X_{ij}, which includes the exogenous variables in X_{ij} and the set of instruments that we discussed in Section 4, the first-stage regression is

$$X_{ij} = \psi Z_{ij} + \alpha_i + \lambda_j + \mu_{ij}. \tag{7}$$

Z_{ij} is a valid set of instruments for X_{ij} as long as $E[Z_{ij}\eta_{ij}] = 0$ and $E[Z_{ij}\gamma_j] = 0$ (and, additionally, Z_{ij} must be correlated with the independent variables for which it instruments). This condition ensures that the instruments are exogenous. If either of these terms is not zero, then the set of instruments is not valid, and the coefficient estimates may be inconsistent.[51]

To be more specific about the direction of the bias in the cross-sectional regressions, we consider the univariate case, and rewrite equation (6) as

$$G_j^{lodd} = \beta_1 x_j + u_j, \tag{8}$$

where x_j corresponds to the level of competition on route j. Assuming that z_j is the distance of route j, and using distance as an instrument for competition in a two-stage least-squares regression, we can write the probability limit of the resulting instrumental variables estimator as:

$$\text{plim}_{n\to\infty}\widetilde{\beta} = \beta_0 + \frac{\text{Cov}(z_j, u_j)}{\text{Cov}(z_j, x_j)}. \tag{9}$$

We argue that $|\frac{\text{Cov}(z_j,u_j)}{\text{Cov}(z_j,x_j)}| > 0$ for the following reasons: As large planes are typically more fuel efficient than small planes, carriers tend to use larger planes on longer routes. Furthermore, due to improved technology in aircraft fuel consumption, it is only in recent years that smaller aircraft have become capable of making long-distance trips. This suggests

[51]We can see this explicitly by writing the expression for the instrumental variables estimator as

$$\beta^{IV} = \beta_0 + (X'P_Z X)^{-1} X' P_Z u,$$

where we have omitted the indices for the route and carrier to simplify notation. β_0 is the true value of the vector of parameters in the population, P_Z is the matrix that projects orthogonally onto the space of Z, and u is the combination of the two error terms, $\gamma_j + \eta_{ij}$. If $\text{plim}_{n\to\infty}(n^{-1}Z'u) \neq 0$, where n is the number of observations, then the instrumental-variables estimator will be inconsistent.

a positive correlation between plane size and route distance, which we see in our data (see Table 7). Price dispersion is also likely related to the size of the plane, since larger planes contain more seats and thus provide an airline with a greater opportunity to implement strategies that create more price dispersion. This suggests a positive relationship between plane size and price dispersion. However, plane size is omitted from the regressions estimated by Borenstein and Rose, so its effect is embedded in the error term, u_{ij}, and implies that the numerator, $\text{Cov}(z_j, u_j)$, is positive. We also find a positive correlation between distance and the degree of competition, which suggests that the denominator, $\text{Cov}(z_j, x_j)$, is negative in the case of $x = \ln HERF$, and positive in the case of $x = \ln N$. Specifically, when we segment routes based on distance, we find that the average number of competitors on a route rises while average concentration falls with distance (see Table 7).[52] While we are not certain about the source of this positive relationship between competition and distance, we believe that it may be due to the fact that monopoly routes tend to be short-distance routes that surround a carrier's hub. Also, longer routes tend to operate out of larger airports, which may be able to handle more competitors. Equation (9) therefore suggests that the sign of the bias is negative in the case of $x = \ln HERF$, and positive in the case of $x = \ln N$.[53]

To address this potential bias in our cross-sectional regressions, we add a measure of plane size (average number of seats for carrier i on route j) as an additional endogenous explanatory variable in our cross-sectional regressions. In essence, we are instrumenting for plane size with distance in an attempt to remove the correlation between distance and the error term. Table 8 reports results from the first quarter of 1993 (the first quarter of our sample) and shows that when we include average seat capacity, $\ln \widehat{ASEATCAP}$, in the regression,[54] its coefficient estimate is positive and significant. More importantly, its inclusion causes the estimated effect of $\ln\widehat{N}$ to fall, and the estimated effect of $\ln\widehat{HERF}$ to rise. In fact, we find that this is true in all 55 quarters for $\ln\widehat{N}$, and in 50 of the 55 quarters for $\ln\widehat{HERF}$.[55] Figure 5 helps illustrate this finding by plotting the coefficient

[52] We also performed univariate regressions of competition on distance and obtained positive and significant coefficient estimates. These results are omitted from the paper but available from the authors upon request.

[53] Borenstein and Rose intended to use distance to instrument for the number of flights on a route, not for the competition variables. However, in a setting with multiple endogenous variables and multiple instruments, it is not possible to assign specific instruments to specific endogenous variables.

[54] The T-100 Segment database contains information on the total number of monthly departures and the total number of available seats. We divide seats by departures to construct a variable corresponding to the average number of available seats per departure, which is our proxy for plane size, $\ln\widehat{ASEATCAP}$.

[55] When $\ln\widehat{ASEATCAP}$ is included as an explanatory variable, the coefficient estimate for $\ln\widehat{N}$ falls to a negative value for 39 of the 55 quarters, and the coefficient estimate for $\ln\widehat{HERF}$ rises to a positive value

estimates associated with $\ln\widehat{HERF}$ before and after the inclusion of $\ln\widehat{ASEATCAP}$ as an explanatory variable. The figure plots estimates for the first quarter of each year, and in all but one of these sample periods the coefficient estimate significantly increases with the inclusion of plane size. This indicates that there is indeed a significant bias induced by the correlations between plane size, distance, and competition. In particular, the bias is in the direction hypothesized.[56]

Our discussion above focused on the effects of distance, which is only one of the time-invariant, route-specific factors that is not controlled for in the cross-sectional regressions. It is certainly possible that other factors besides distance are also affecting the cross-sectional estimates of the effects of competition on price dispersion. Our results from the panel analysis, in which we are able to control for all time-invariant effects, provide some support for this view. Overall, the results from our cross-sectional estimation controlling for plane size, combined with our more conclusive findings from the fixed-effects, panel analysis, lead us to support the view that the monopoly effect is dominant.

7 Conclusion

In this study, we update and extend the cross-sectional analysis of Borenstein and Rose (1994), and we also perform a panel analysis where we use fixed-effects estimation that controls for all time-invariant, carrier-route-specific factors. This enables us to identify the effects of changes in the competitive structure of a route on price dispersion over time. In our estimation of the cross-section, we were able to replicate Borenstein and Rose's results for most of the quarters in our sample. However, the results of our panel analysis—where any bias induced by time-invariant, route-specific effects has been removed—contradict their results. Our results lend support to the monopoly effect, showing that a decrease in market concentration (that is, an increase in competition) over time along a route results in a decrease in price dispersion. Our findings show that Borenstein and Rose's cross-sectional

for 26 of the 55 quarters.

[56] We also performed regressions in which we included distance as a right-hand-side variable. In most of the time periods, distance entered with a positive and significant coefficient estimate, and its inclusion subsequently lowered the estimate for $\ln\widehat{N}$ and raised the estimate for $\ln\widehat{HERF}$. We also found a significant reduction in the J-statistic—which tests the over-identification of the instrument set—when we removed distance from the instrument set and also when we left distance in the instrument set but included average seat capacity in the regression under standard IV estimation. For example, in 1993:Q1 (under model 1), the J-statistic falls from 256.8 to 87.4 when we remove distance from the instrument set. It falls to 11.7 when we put distance back into the instrument set but include $\ln\widehat{ASEATCAP}$ as an explanatory variable. These results suggest that distance is highly correlated with the error term when plane size is not included in the regression.

estimates for route concentration suffer from a downward bias, due to the omission of plane size as an explanatory variable and the inclusion of distance in the instrument set.

We find that an increase in the number of carriers on a route significantly reduces price dispersion on routes that we identify as having a heterogeneous mixture of business travelers and leisure travelers. Specifically, an increase over time in the number of carriers on these routes lowers the prices at the top of the distribution to a greater extent than it lowers prices at the bottom of the distribution, resulting in a decline in overall price dispersion. On routes to leisure destinations, where we believe the consumer base is more homogeneous, the effects of competition on price dispersion are largely insignificant. On leisure routes, following an increase in the number of competitors, prices in the top and bottom ends of the distribution fall by similar amounts, resulting in smaller, and less significant changes in price dispersion. This suggests that prices charged to business travelers are more affected by increased competition than prices charged to leisure travelers. Furthermore, our study indicates that competition from LCCs and regional carriers has a stronger negative effect on price dispersion than competition from other legacy carriers. This signifies that business travelers have benefited from the emergence of LCCs to an even greater extent than leisure travelers.

Overall, we show that competition decreases the ability of a carrier to charge a high price to price-inelastic consumers relative to the price it charges price-elastic consumers, eroding its ability to segment markets. It is clear that this erosion is due to a loss of market share that typically accompanies the entry of a new competitor. Thus, we find striking evidence of the monopoly effect on prices, which supports the traditional textbook treatment of the relationship between competition and price discrimination.

References

[1] Alam, I., Ross, L., Sickles, R., 2001. Time Series Analysis of Strategic Pricing Behavior in the US Airline Industry. *Journal of Productivity Analysis* 16: 49–62.

[2] Berry, S., 1990. Airport Presence as Product Differentiation. *American Economic Review* 80: 394–399.

[3] Berry, S., 1992. Estimation of a Model of Entry in the Airline Industry. *Econometrica* 60(4): 889–917.

[4] Borenstein, S., 1985. Price Discrimination in Free-Entry Markets. *The RAND Journal of Economics* 16(3): 380–397.

[5] Borenstein, S., 1989. Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry. *The RAND Journal of Economics* 20(3): 344–365.

[6] Borenstein, S., Rose, N., 1994. Competition and Price Dispersion in the U.S. Airline Industry. *The Journal of Political Economy* 102(4): 653–683.

[7] Borenstein, S., Rose, N., 1995. Bankruptcy and Pricing Behavior in U.S. Airline Markets. *The American Economic Review, Papers and Proceedings* 85(2): 397–402.

[8] Brueckner, J., Spiller, P., 1991. Competition and Mergers in Airline Networks. *International Journal of Industrial Organization* 9: 323–342.

[9] Brueckner, J., Dyer, N., Spiller, P., 1992. Fare determination in airline hub-and-spoke networks. *The RAND Journal of Economics* 23(3): 309–333.

[10] Busse, M., and Rysman, M., 2005. Competition and Price Discrimination in Yellow Pages Advertising. *The RAND Journal of Economics* 36(2): 378–390.

[11] Cowell, F. A., 1995. *Measuring Inequality* (second edition). Hemel Hempstead, England: Harvester Wheatsheaf.

[12] Dana, J., 1999. Equilibrium Price Dispersion Under Demand Uncertainty: The Roles of Costly Capacity and Market Structure. *The RAND Journal of Economics* 30(4): 632–660.

[13] Eden, B., 1990. Marginal Cost Pricing When Spot Markets Are Complete. *The Journal of Political Economy* 98(6): 1293–1306.

[14] Evans, W., Kessides, I., 1993. Localized Market Power in the U.S. Airline Industry. *The Review of Economic and Statistics* 75(1): 66–75.

[15] Gale, I., 1993. Price Dispersion in a Market with Advance-Purchases. *Review of Industrial Organization* 8: 451–464.

[16] Goolsbee, A. and Syverson, C., 2005. How do Incumbents Respond to the Threat of Entry? Evidence from Major Airlines. NBER Working Paper 11072.

[17] Hayes, K., Ross, L., 1998. Is Airline Price Dispersion the Result of Careful Planning or Competitive Forces? *Review of Industrial Organization* 13: 523–542.

[18] Holmes, T., 1989. The Effects of Third-Degree Price Discrimination in Oligopoly. *The American Economic Review* 79(1): 244–250.

[19] Ito, H., Lee, D., 2003a. Low Cost Carrier Growth in the U.S. Airline Industry: Past, Present, and Future. Brown University Working Paper, No. 2003–12.

[20] Ito, H., Lee, D., 2003b. Incumbent Responses to Lower Cost Entry: Evidence from the U.S. Airline Industry. Brown University Working Paper, No. 2003–22.

[21] Lott, J., Roberts, R., 1991. A Guide to the Pitfalls of Identifying Price Discrimination. *Economic Inquiry* 29(1): 14–23.

[22] Prescott, E., 1975. Efficiency of the Natural Rate. *The Journal of Political Economy* 83(6): 1229–1236.

[23] Stavins, J., 1996. Price Discrimination in the Airline Market: The Effect of Market Concentration. *Review of Economics and Statistics* 83(1), 200–202.

[24] Varian, H., 1989. Price discrimination. *Handbook of Industrial Organization*, edited by R. Schmalensee and R. Willig. Amsterdam: North-Holland.

[25] Yang, J., Liu A., 2003. Frequent Flyer Program: A Case Study of China Airline's Marketing Initiative-Dynasty Flyer Program. *Tourism Management* 24(5): 587–595.

Table 1: Summary of Competition and Price Dispersion Statistics

1993 Quarter 1	(1) All Routes	(2) Big-City Routes	(3) Leisure Routes
Average Gini coefficient	0.22	0.25	0.20
Average number of competitors per route	3.02	3.77	2.96
Inverse of the Herfindahl index (Average)	1.56	1.74	1.74
2006 Quarter 3	All Routes	Big-City Routes	Leisure Routes
Average Gini coefficient	0.22	0.23	0.18
Average number of competitors per route	2.61	3.12	3.25
Inverse of the Herfindahl index (Average)	1.57	1.73	1.78

Table 2: Leisure-Routes Sample

Airport	Airport City and State	$\frac{Accommodation}{Non-FarmEarnings}$ (median)	Percentile	Included in Leisure Sample?
ACY	Atlantic City, NJ	0.2612		yes
LAS	Las Vegas, NV	0.1702		yes
OGG	Kahului, HI	0.1665		yes
MKK	Kaunakakai, HI	0.1665		yes
JHM	Lahaina, HI	0.1665		yes
LNY	Lanai City, HI	0.1665		yes
LIH	Kapaa, HI	0.1293		yes
ITO	Hilo, HI	0.0974		yes
KOA	Kona, HI	0.0974	95	yes
RNO	Reno, NV	0.0734		yes
EGE	Eagle County, CO	0.0605		yes
MYR	Myrtle Beach, SC	0.0539		yes
FLG	Flagstaff, AZ	0.0376		yes
MCO	Orlando, FL	0.0305		yes
HNL	Honolulu, HI	0.0283		yes
SAV	Hilton Head, SC	0.0276		yes
BZN	Bozeman, MT	0.0275		yes
DSW	Naples, FL	0.0264	90	yes
SAF	Santa Fe, NM	0.0245		yes
MEM	Memphis, TN	0.0214		no
KTN	Ketchikan, AK	0.0199		yes
PFN	Panama City, FL	0.0192		yes
SBA	Santa Barbara, CA	0.0161		yes
MSY	New Orleans, LA	0.0148		yes
CHS	Charleston, SC	0.0139		yes
RAP	Rapid City, SD	0.0132	85	yes
SJU	San Juan, PR	N/A		yes
STX	St. Croix, USVI	N/A		yes
STT	St. Thomas, USVI	N/A		yes

Table 3: Big-City Routes Sample

Airport	**Metropolitan Area**	Population	Included in Big-City Sample?
JFK, EWR, LGA	New York/Northern New Jersey/Long Island	**18,747,320**	yes
LAX, SNA, BUR	Los Angeles/Long Beach/Santa Ana	**12,923,547**	yes
ORD, MDW	Chicago/Naperville/Joliet	**9,443,356**	yes
PHL	Philadelphia/Camden/Wilmington	**5,823,233**	yes
DAL, DFW	Dallas/Fort Worth/Arlington	**5,819,475**	yes
FLL, MIA	Miami/Fort Lauderdale/Pompano Beach	**5,422,200**	no
IAH	Houston/Sugar Land/Baytown	**5,280,077**	yes
DCA, IAD	Washington/Arlington/Alexandria	**5,214,666**	yes
ATL	Atlanta/Sandy Springs/Marietta	**4,917,717**	yes
DTW	Detroit/Warren/Livonia	**4,488,335**	yes
BOS	Boston/Cambridge/Quincy	**4,411,835**	yes
SFO, OAK, SJC	San Francisco/Oakland/Fremont	**4,152,688**	yes
ONT	Riverside/San Bernardino/Ontario	**3,909,954**	yes
PHX	Phoenix/Mesa/Scottsdale	**3,865,077**	yes
SEA	Seattle/Tacoma/Bellevue	**3,203,314**	yes
MSP	Minneapolis/St. Paul/Bloomington	**3,142,779**	yes
SAN	San Diego/Carlsbad/San Marcos	**2,933,462**	no
STL	St. Louis	**2,778,518**	yes
BWI	Baltimore/Towson	**2,655,675**	yes
TPA	Tampa/St. Petersburg/Clearwater	**2,647,658**	no
PIT	Pittsburgh	**2,386,074**	yes
DEN	Denver/Aurora	**2,359,994**	yes
CLE	Cleveland/Elyria/Mentor	**2,126,318**	yes
PDX	Portland/Vancouver/Beaverton	**2,095,861**	yes
CVG	Cincinnati/Middletown	**2,070,441**	yes
SMF	Sacramento/Arden/Arcade/Roseville	**2,042,283**	yes
MCI	Kansas City	**1,947,694**	yes
MCO	Orlando/Kissimmee	**1,933,255**	no
SAT	San Antonio	**1,889,797**	yes
SJC	San Jose/Sunnyvale/Santa Clara	**1,754,988**	yes

Table 4: Cross-Sectional Estimates: Dependent Variable is G_{ij}^{lodd}

	1993:Q1		1999:Q1		2005:Q1	
	(1)	(2)	(1)	(2)	(1)	(2)
$\ln \widehat{HERF}$	-0.081**		-0.201***		-0.079***	
	(0.036)		(0.038)		(0.029)	
$\ln MK\widehat{TSH}ARE$	0.040***		0.080***		0.087***	
	(0.009)		(0.011)		(0.011)	
$\ln \widehat{N}$		0.131***		0.179***		-0.014
		(0.026)		(0.029)		(0.026)
$\ln F\widehat{LTTO}T$	-0.037***	-0.075***	-0.035***	-0.081***	0.015	-0.004
	(0.011)	(0.011)	(0.011)	(0.012)	(0.010)	(0.011)
$\ln TOURIST$	-0.001	-0.017*	-0.030***	-0.045***	-0.024***	-0.026***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)
HUB	0.031*	0.101***	0.075***	0.215***	0.033	0.118***
	(0.018)	(0.017)	(0.023)	(0.020)	(0.021)	(0.020)
$SMALL$	-0.109***	-0.058**	-0.127***	-0.097***	-0.020	-0.026
	(0.024)	(0.026)	(0.027)	(0.029)	(0.025)	(0.027)
Observations	1744	1744	1556	1556	1280	1280

Notes: All regressions include carrier-specific dummies. Route-specific effects are considered random in these regressions. Robust standard errors are in parentheses. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table 5: Panel Estimates: Dependent Variable is G_{ijt}^{lodd}

	All Routes			Big-City Routes			Leisure Routes		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\ln \widehat{HERF}$	0.063**			0.134***			0.036		
	(0.016)			(0.031)			(0.028)		
$\ln \widehat{N}$		-0.059***			-0.138***			-0.029	
		(0.018)			(0.034)			(0.039)	
$LCCdum$			-0.021***			-0.054***			-0.003
			(0.007)			(0.011)			(0.015)
$SWdum$			-0.044***			-0.036*			-0.045*
			(0.010)			(0.019)			(0.025)
$LEGACYdum$			-0.006			-0.017			0.008
			(0.006)			(0.011)			(0.012)
$REGIONALdum$			-0.009			-0.033***			-0.006
			(0.008)			(0.012)			(0.021)
$LCCthreat$			0.009**			0.017**			-0.003
			(0.004)			(0.007)			(0.011)
$SWthreat$			-0.013			-0.003			-0.034
			(0.008)			(0.014)			(0.025)
Observations	82855	82855	82861	27030	27030	27031	15666	15666	15667

Notes: All regressions include route-carrier-specific dummies, time dummies, quarter dummies, and a dummy variable indicating whether the carrier is in bankruptcy. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table 6: Panel Estimates: Dependent Variable is Log of 90th- and 10th-Percentile

	Big City Routes						Leisure Routes					
	$\ln P90$			$\ln P10$			$\ln P90$			$\ln P10$		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
$\ln \widehat{HERF}$	0.390***			0.188***			0.185***			0.139***		
	(0.046)			(0.031)			(0.036)			(0.035)		
$\ln \widehat{N}$		-0.597***			-0.282***			-0.306***			-0.224***	
		(0.070)			(0.043)			(0.061)			(0.058)	
$LCCdum$			-0.246***			-0.109***			-0.124***			-0.067***
			(0.026)			(0.015)			(0.024)			(0.019)
$SWdum$			-0.211***			-0.111***			-0.239***			-0.162***
			(0.049)			(0.023)			(0.032)			(0.023)
$LEGACYdum$			-0.076***			-0.048***			-0.014			-0.030**
			(0.020)			(0.015)			(0.015)			(0.014)
$REGIONALdum$			-0.127***			-0.050***			-0.059**			-0.024
			(0.022)			(0.016)			(0.025)			(0.017)
$LCCthreat$			-0.014			-0.046***			-0.034**			-0.013
			(0.017)			(0.010)			(0.017)			(0.014)
$SWthreat$			-0.065*			-0.051***			-0.182***			-0.118***
			(0.035)			(0.017)			(0.033)			(0.027)
Observations	27030	27030	27031	27030	27030	27031	15666	15666	15667	15666	15666	15667

Notes: All regressions include route-carrier-specific dummies, time dummies, quarter dummies, and a dummy variable indicating whether the carrier is in bankruptcy. Standard errors are in parentheses and are clustered by route to account for both autocorrelation and correlation between carriers on the same route. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Table 7: Distance, Competition, and Plane Size

				$\bar{N}$	$H\bar{E}RF$	$ASEA\bar{T}CAP$
	distance	$<$	450	2.49	.78	128
$450 \leq$	distance	$<$	818	2.23	.77	128
$818 \leq$	distance	$<$	1240	2.85	.73	139
$1240 \leq$	distance			3.16	.72	167

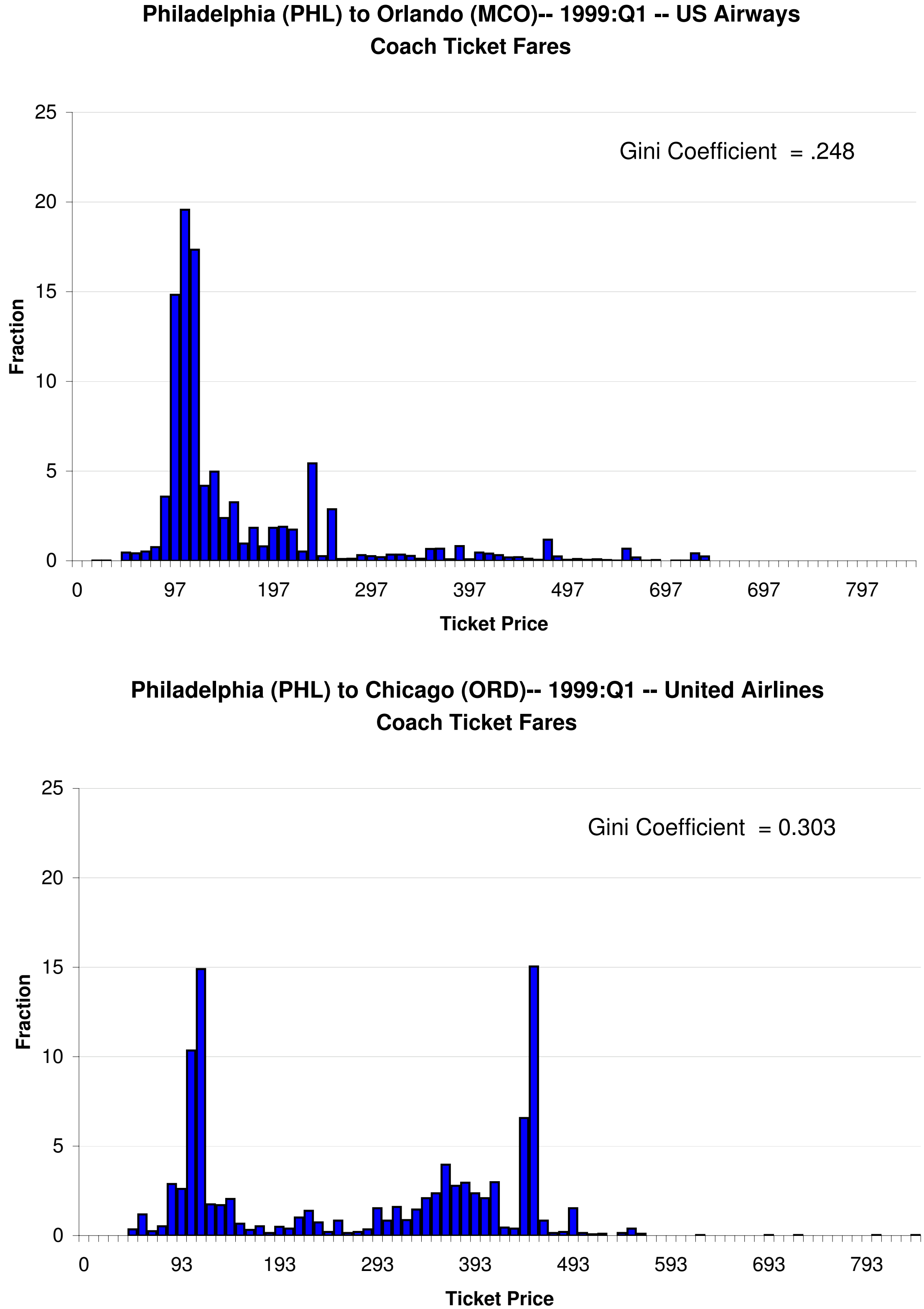
Notes: Each row represents a quartile of our sample, based on distance. $\bar{N}$ is the average number of competitors, $H\bar{E}RF$ is the average Herfindahl, and $ASEA\bar{T}CAP$ is the average seat capacity of routes in each quartile.

Table 8: 1993:Q1 Cross-Section Robustness Estimates: Dependent Variable is G_{ij}^{lodd}

	(1)	(2)	(3)	(4)
$\ln \widehat{HERF}$	-0.081**		0.019	
	(0.036)		(0.047)	
$\ln \widehat{MKTSHARE}$	0.040***		-0.012	
	(0.009)		(0.012)	
$\ln \widehat{N}$		0.131***		0.019
		(0.026)		(0.033)
$\ln \widehat{FLTTOT}$	-0.037***	-0.075***	-0.084***	-0.083***
	(0.011)	(0.011)	(0.014)	(0.013)
$\ln TOURIST$	-0.001	-0.017*	-0.085***	-0.081***
	(0.010)	(0.010)	(0.014)	(0.013)
HUB	0.031*	0.101***	0.068***	0.062***
	(0.018)	(0.017)	(0.022)	(0.020)
$SMALL$	-0.109***	-0.058**	-0.014	-0.014
	(0.024)	(0.026)	(0.032)	(0.031)
$\ln \widehat{ASEATCAP}$			0.963***	0.921***
			(0.083)	(0.078)
Observations	1744	1744	1743	1743

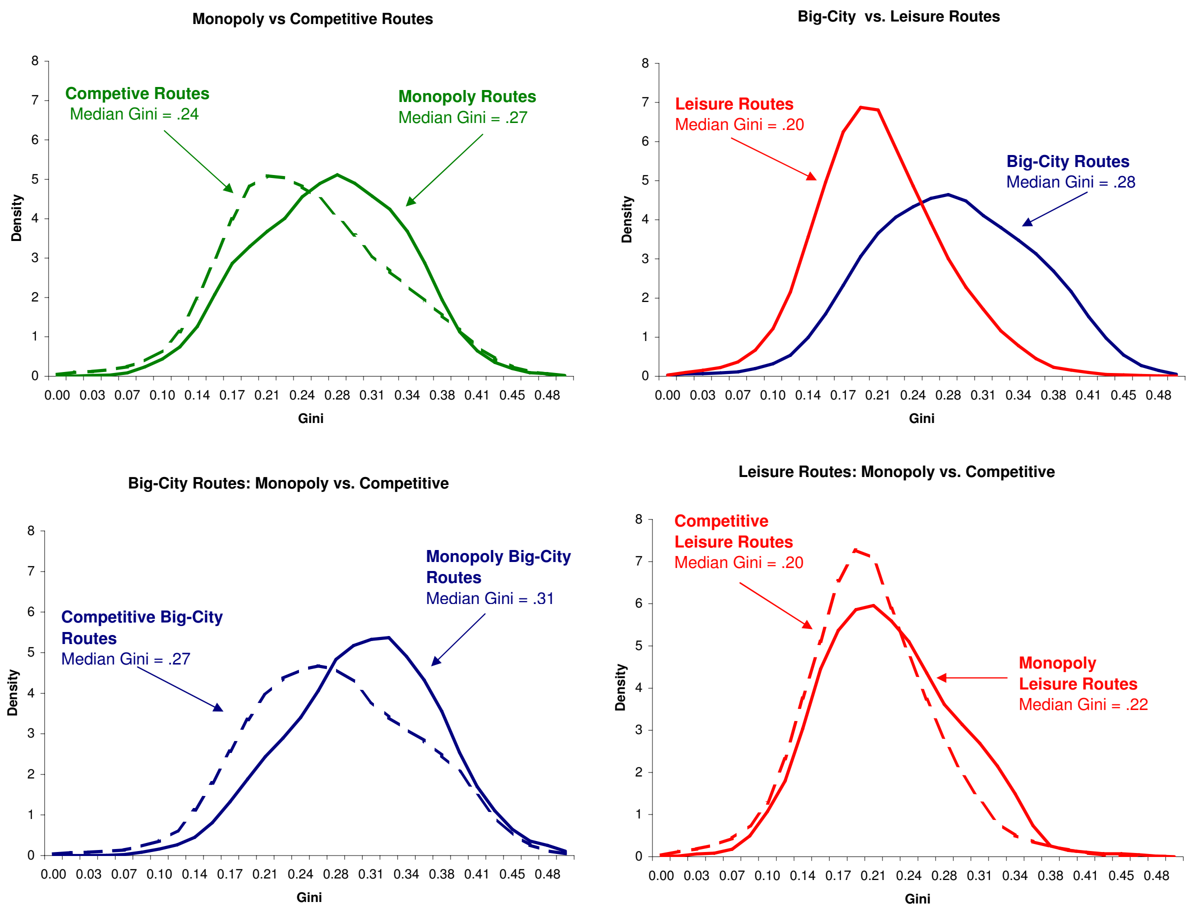
Notes: All regressions include carrier-specific dummies. Route-specific effects are considered random. Robust standard errors are in parentheses. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

Figure 1: Example Price Distributions



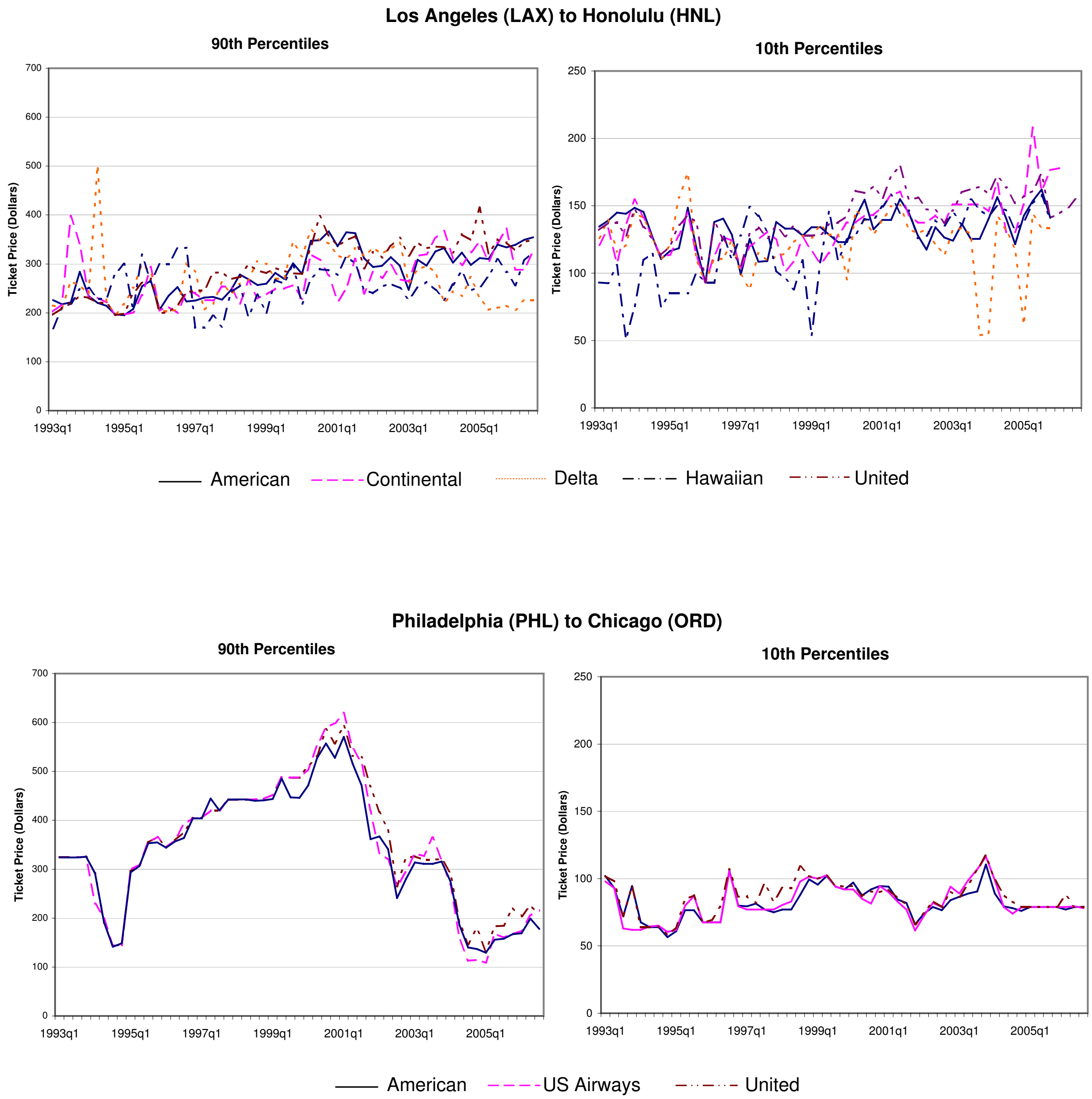
Notes: This figure displays histograms of coach-class fares for US Airways' Philadelphia-to-Orlando route and United Airlines' Philadelphia-to-Chicago route during the first quarter of 1999. Prices are in nominal U.S. dollars and are computed as directional fares (round-trip fares are divided by two).

Figure 2: PDFs of Gini Coefficients of Ticket Prices: 1993:Q1 to 2006:Q3



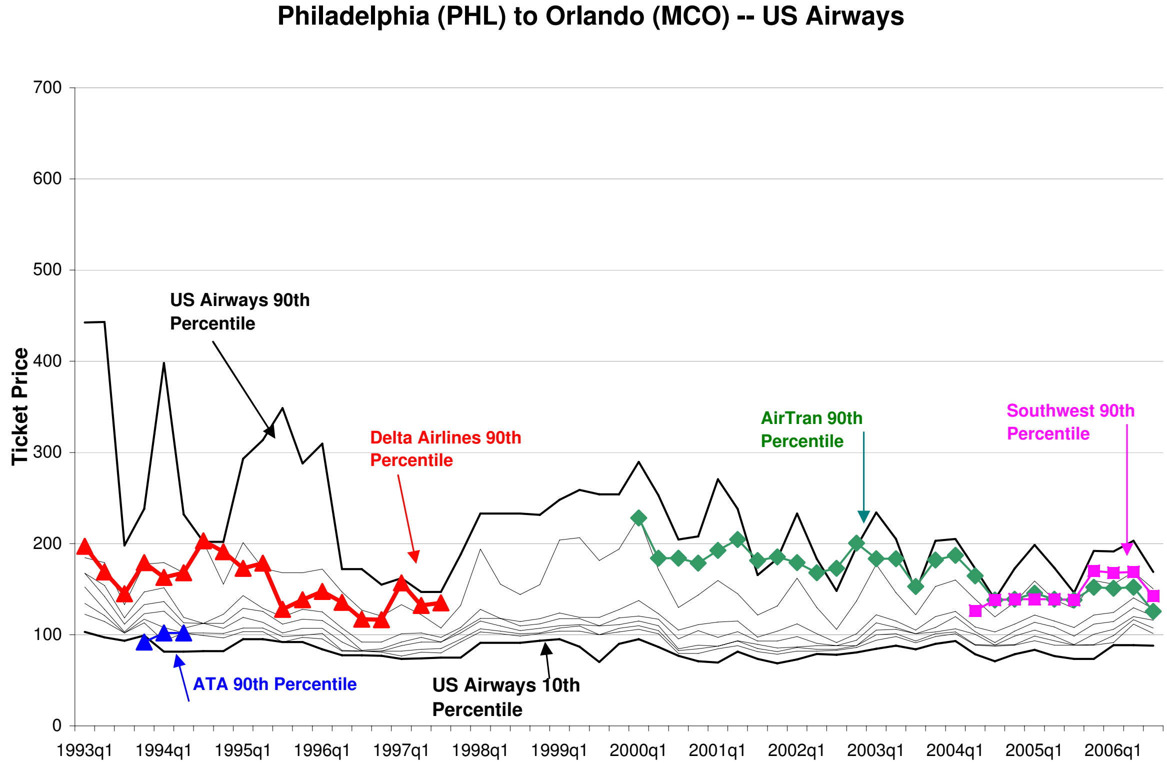
Notes: This figure shows probability density functions of the Gini coefficient of domestic coach-class fares after conditioning the sample on certain restrictions. Conditional restrictions, along with the median Gini coefficient in the specified sample, are depicted next to each density function. Monopoly routes are those in which the average market share over the sample period is greater than 0.95 for a single carrier.

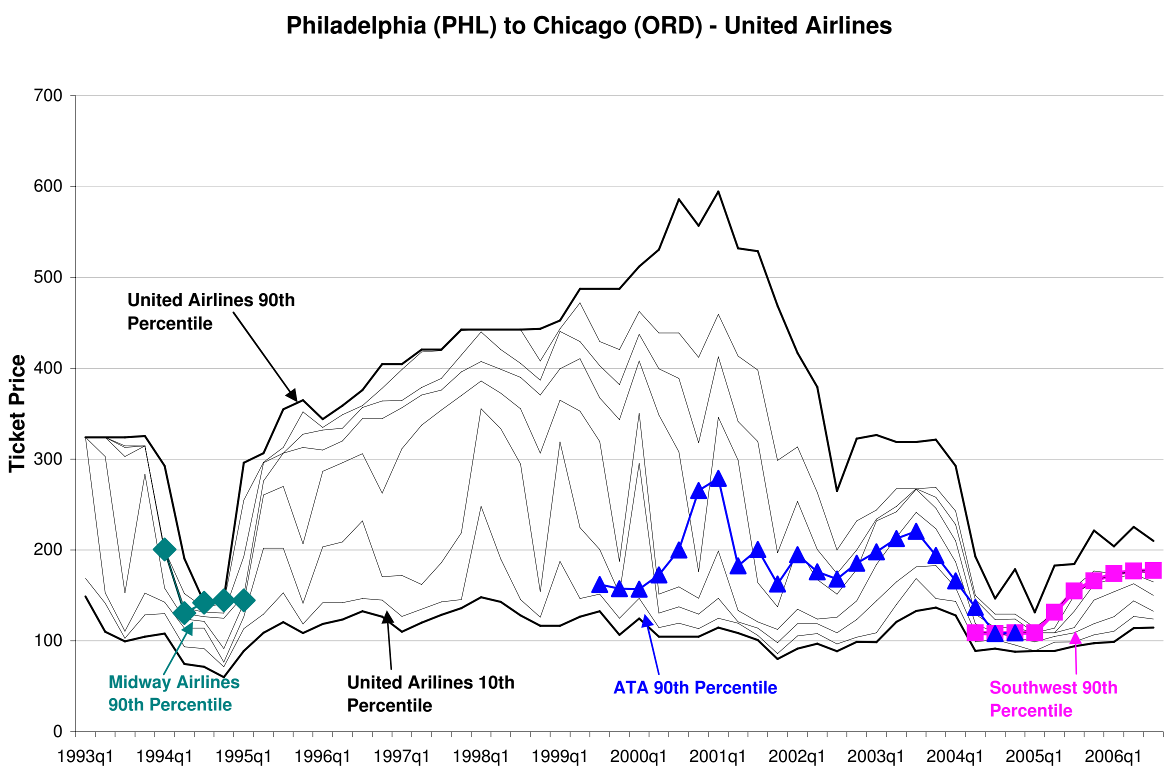
Figure 3: Pricing Dynamics - Incumbents



Notes: This figure compares the 90th and 10th percentiles of coach-class fares between carriers operating on the specified routes over the entire sample period. The above two panels represent a "leisure route" while the bottom two panels represent a "big-city route."

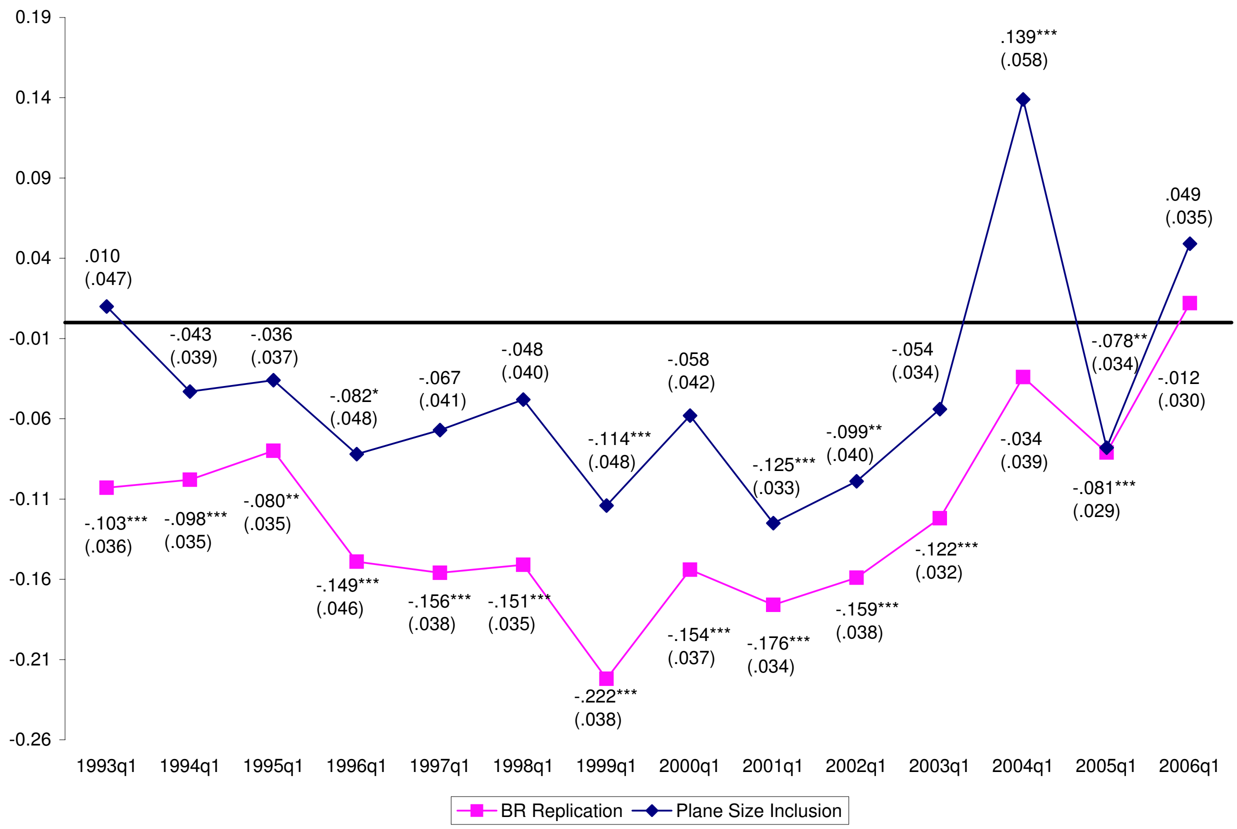
Figure 4: Pricing Dynamics - Entry and Exit





Notes: This figure shows the entry and exit of carriers into two specified routes. Depicted are the 90th percentiles of the entrants, and the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of the incumbent carrier.

Figure 5: Coefficient Estimates for $\ln\widehat{HERF}$



Notes: This figure shows estimates of the coefficient on $\ln\widehat{HERF}$ under equation (1) for quarter 1 of each year in our sample. Squares indicate estimates of our Borenstein and Rose replication, while diamonds indicate estimates of the model including average seat capacity as an endogenous explanatory variable. Robust standard errors are in parenthesis. One, two, and three asterisks indicate significance at the 10-percent, 5-percent, or 1-percent significance level, respectively.

A Data Construction

In this appendix, we discuss our methods and assumptions involved in constructing our panel of airline-route ticket observations from the DB1B and T-100 Domestic Segment databases maintained by the BTS at their online website, Transtats. In order to provide readers with an idea of the quantitative significance of each of our assumptions, we use the first quarter of 1993 (the first quarter of our data) as an example.

There are three different sub-components to the DB1B data set. They are market data, coupon data, and ticket data, and we combine variables from all three.[57] To maintain consistency with Borenstein and Rose (1994), we restrict our sample to only the traditional, legacy carriers. The number of legacy carriers fluctuates between 8 and 9 over the course of our sample, as Trans World Airlines was bought out by American Airlines in April of 2001. In the first quarter of 1993, we lose 71,475 tickets (4.84 percent of the total number of tickets) with this restriction.

We use only domestic, coach-class itineraries and keep only tickets containing non-stop (direct) flights. In 1993:Q1, we identified 45.9 percent of the tickets issued by legacy airlines as direct flights (644,280 observations out of 1,403,624 total observations). However, this number is inflated, as it includes round-trip tickets that include a direct flight in one direction, but a non-direct flight in the other direction. Therefore, in an effort to restrict our sample to "pure" direct flights, we drop all of these tickets (an additional 205,408 tickets in the first quarter of 1993). The BTS includes a variable that describes the reliability of each ticket price ("dollar cred"). The variable takes on a value of 0 if the fare is of questionable magnitude, based on a set of limits defined by the BTS, and it takes a value of 1 if it is credible. We drop all observations for which this variable takes a value of zero. In 1993:Q1, approximately 0.95 percent of direct tickets issued by legacy carriers had a 0 value for "dollar cred" (4,147 tickets).

The DB1B provides limited information regarding the fare class of each ticket. We eliminated all first-class and business-class itineraries. In 1993:Q1, 4.00 percent of direct, legacy tickets in the DB1B were identified as either business class or first class (17,383 tickets). We also eliminated any tickets that we believe to be frequent-flyer tickets, using the following method. First, we dropped all fares coded as 0 (13,642 tickets in quarter 1 of 1993). Next, we dropped all fares that were less than or equal to \$20 (\$10 for one-way tickets), which amounted to 8,448 additional tickets in the first quarter of 1993. Combined, this implies that 5.29 percent of coach-class, direct, legacy tickets were frequent-flyer tickets

[57]For further reference, see the BTS's website http://www.transtats.bts.gov.

in the first quarter of 1993. This left us with 395,252 tickets from the DB1B in the first quarter of 1993, covering 7,273 distinct directional routes.

After merging all 55 quarters of the DB1B data with supplemental data from the T-100 Segment data, we were left with 195,044 airline-route observations (an example of an airline-route observation is a United Airlines flight from Boston to Orlando in the fourth quarter of 2000), which encompass 6,272 distinct directional routes. The merge between the DB1B and T-100 Segment databases was not exact (around 50 percent matched). First, since the DB1B does not provide complete coverage for all airlines and routes, there are a number of direct routes in the T-100 data that we do not find in the DB1B (especially low-volume routes). Second, the DB1B does not allow us to distinguish between a non-stop, direct ticket and a ticket that involves a stop without a plane change. For example, if a passenger takes a flight from Boston to Orlando that stops in Atlanta but does not involve a plane change, his itinerary will look identical to that of a passenger who flies from Boston to Orlando without any stops. For this reason, we identified some airline routes as direct in the DB1B, that are not truly direct and therefore do not have segment information in the T-100 data. While we lose many airline-route observations during the merge as a result, we believe that this merge actually provides a nice filter, since we would ideally like to use only non-stop, direct flights. Thus, by merging data between the DB1B and the T-100, we likely eliminate a large proportion of flights that are not really direct in nature.

In an effort to eliminate possible coding errors, we drop certain airline-route observations from the data that we believe do not have adequate coverage to calculate reliable price dispersion statistics. We drop any airline-route observation that does not have at least 100 passengers in the DB1B. Furthermore, for each airline route observation, we calculate the average number of passengers over time in both the DB1B and the T-100 Segment databases. If the number of passengers on an airline route in a given quarter falls below 25 percent of its mean over time in one of the databases, but not in the other, then we drop the observation from our data, on the basis that its value is most likely measurement error. However, if the number of passengers on an airline route in a given quarter falls below 25 percent of its mean in both the DB1B and the T-100 Segment databases, then we keep the observation in our data. This leaves us with 160,751 airline-route observations. which cover 3,733 distinct directional routes.

Finally, we addressed the issue of "double counting." Since we defined a route as a directional trip in our data, any round-trip ticket would count twice. For example, a round-trip fare from Boston to San Francisco would appear twice in the data — once as BOS-SFO and once as SFO-BOS. Since this would have no effect on the consistency of our estimates, but a significant effect on the size of our standard errors, we chose to drop one of the directions. Of course, the drawback of this assumption is that some one-way fares were dropped from the data as a result. In our judgment, the first issue outweighed the second issue. Dropping one of the directions decreases our sample to 82,855 observations, which cover 2,752 distinct routes.

B Variable Definitions

- G_{ijt}^{lodd} - The Gini log-odds ratio, given by $G_{ijt}^{lodd} = \frac{ln(G_{ijt})}{1-ln(G_{ijt})}$, where G_{ijt} is the Gini coefficient of carrier i's price distribution on route j in period t, calculated using data from DB1B.
- $\ln P(k)_{ijt}$ - The logarithm of the kth price percentile of carrier i on route j in period t, obtained from the DB1B.
- $\ln MKTSHARE_{ijt}$ - The logarithm of the share of total passengers originating on route j operated by carrier i in period t, calculated from the DB1B.
- $\ln HERF_{jt}$ - The logarithm of the Herfindahl index of route j in period t, calculated using passenger shares obtained from the DB1B.
- $\ln N_{jt}$ - The logarithm of the total number of carriers operating on route j in period t, obtained from DB1B.
- $\ln FLTTOT_{jt}$ - The logarithm of the total number of departures performed on route j in period t, obtained from the T-100 Domestic Segment Databank.
- HUB_{ij} - A dummy variable indicating whether either the origin or destination of route j is a hub airport of carrier i.
- $SMALL_j$ - A dummy variable indicating if both the origin and the destination airport are not in our list of big cities.
- $\ln TOURIST_j$ - The logarithm of the maximum of the ratio of accommodation earnings to total non-farm earnings for the origin and destination cities on route j, obtained from the Bureau of Economic Analysis.
- $\ln ASEATCAP_{ij}$ - The logarithm of average seat capacity (total available seats divided by total number of departures) on route j by carrier i obtained from the T-100 Domestic Segment Databank.
- $SWdum_{jt}$ - A dummy variable indicating whether Southwest Airlines operates on the same city pair as the observed carrier i in period t.
- $SWthreat_{jt}$ - A dummy variable indicating whether Southwest Airlines operates out of both cities of route j in period t, but does not operate the route between them.
- $LCCdum_{jt}$ - A dummy variable indicating whether a specific low-cost carrier (not including Southwest Airlines) operates on the same city pair as the observed carrier i in period t.

- $LCCthreat_{jt}$ - A dummy variable indicating whether a specific low-cost carrier (not including Southwest Airlines) operates out of both cities of route j in period t, but does not operate the route between them.

- $LEGACYdum_{jt}$ - A dummy variable indicating whether another legacy carrier operates on the same city pair as the observed carrier i in period t.

- $REGIONALdum_{jt}$ - A dummy variable indicating whether a regional carrier operates on the same city pair as the observed carrier i in period t.

Instruments

- $\ln DISTANCE_j$ - The logarithm of nonstop distance in miles between endpoint airports of route j.

- $AMEANPOP$ - The arithmetic mean of the metropolitan population of endpoint cities taken from the 2000 U.S. Census.

- $GMEANPOP$ - The geometric mean of the metropolitan population of endpoint cities taken from the 2000 U.S. Census.

- $\ln PASSRTE_{jt}$ - The logarithm of total enplaned passengers on route j in period t from the T-100 Domestic Segment Databank.

- $IRUTHERF$ - This instrument is identical to one used by Borenstein and Rose (1994). This variable is the square of the fitted value for $MKTSHARE_{ijt}$ from its first-stage regression, plus the rescaled sum of the squares of all other carrier's shares. See Borenstein and Rose (1994) for a more detailed explanation. It is equal to $\widehat{MKTSHARE}_{ijt}^{\,2} + \frac{HERF_{jt} - MKTSHARE_{ijt}^2}{(1-MKTSHARE_{ijt})^2} * (1 - \widehat{MKTSHARE}_{ijt})^2$.

- $GENSP$ - $\frac{\sqrt{ENP_{j1} * ENP_{j2}}}{\sum_k \sqrt{ENP_{k1} * ENP_{k2}}}$, where k indexes all airlines, j is the observed airline, and ENP_{k1} and ENP_{k2} are airline k's average quarterly enplanements at the two endpoint airports. This instrument is similar to one used by Borenstein and Rose (1994), with the difference being that Borenstein and Rose use average daily enplanements, while we use average quarterly enplanements, as a result of data availability. Data on enplanements were obtained from the T-100 Domestic Segment Databank.

C Expanded Summary Statistics

1993 Quarter 1	(1) All Routes	(2) Big-City Routes	(3) Leisure Routes
Average number of competitors per route	3.02	3.77	2.96
Inverse of the Herfindahl index	1.56	1.74	1.74
Average fare (nominal $)	197	220	163
Median fare (nominal $)	181	192	143
Average route distance (miles)	832	976	1,021
Average seat capacity (seats)	143	148	165
Average departures per route	300	374	224
Percent of oneway tickets	9.4	8.7	12.1
Inter-quartile range	127	167	65
Coefficient of Variation	0.45	0.49	0.43
Log Variance	0.20	0.27	0.15
Average Gini coefficient	0.22	0.25	0.20
Average Atkinson index	0.05	0.06	0.04
Average Theil index	0.67	1.10	0.75
Number of routes	869	276	161
Number of airlines	9	9	9

2006 Quarter 3	All Routes	Big-City Routes	Leisure Routes
Average number of competitors per route	2.61	3.12	3.25
Inverse of the Herfindahl index	1.57	1.73	1.78
Average fare (nominal $)	199	204	188
Median fare (nominal $)	174	173	173
Average route distance (miles)	1150	1,194	1,513
Average seat capacity (seats)	148	144	167
Average departures per route	321	443	236
Percent of oneway tickets	12.6	11.7	11.8
Inter-quartile range	93	105	66
Coefficient of Variation	0.46	0.49	0.41
Log Variance	0.2	0.22	0.16
Average Gini coefficient	0.22	0.23	0.18
Average Atkinson index	0.05	0.05	0.04
Average Theil index	0.67	0.75	0.73
Number of routes	783	360	172
Number of airlines	8	8	8

D Borenstein and Rose Variable Comparison

Variable	Borenstein and Rose (1994)	Cross-Section	Panel	Reason for difference
Price Dispersion	ln$GINI$ – The log of the Gini coefficient)	G^{lodd}	G^{lodd}	See Section 4.1.1.
Market Share	ln$FLTSHARE$ – The proportion of weekly total flights on route j	ln$MKTSHARE$	Omitted	See footnotes 24 and 34 and Section 5.1.
Concentration	ln$FLTHERF$ – The herfindahl index based on $FLTSHARE$ as the measure of market share.	ln$HERF$	ln$HERF$	See footnotes 24 and 34 and Section 5.1.
Flight Density	ln$FLTTOT$	ln$FLTTOT$	Omitted	See Section 5.1.
Population Heterogeniety	ln$TOURIST$	ln$TOURIST$	Omitted	See footnote 43.
Airport Dominance	ln$ENDOMO$ – Weighted average of each carrier's share of passengers at each endpoint.	HUB	Omitted	We felt HUB was a more exogenous variable.
Uncongested Airport	$DUMAPT$ – Equal to one if route not included on their list of 24 congested airports.	$SMALL$	Omitted	See footnote 43.
Shadow Flight Capacity Cost	ln$SDCAPFLT$ – The standard deviation of cubed weekly fleet utilization rates.	Omitted	Omitted	Unavailable data.
Shadow Airport Capacity Cost	ln$SDCAPAPT$ – The standard deviation of cubed weekly airport capacity utilization.	Omitted	Omitted	Unavailable data.
Intraquarter Fare Variation	$INQGINI$ – Intertemporal Gini coefficient of the lowest fare	Omitted	Omitted	Unavailable data.