Revisiting The Relationship Between Competition And Price Dispersion*

Yihui Fu †
Georgetown University

Qian Wang §
University of Maryland

November 29, 2022

Abstract

Using a novel granular data covering price and booking information of flights in China, we test the theory that predicts how the competition will affect price dispersion. We complement the past price dispersion studies by making two contributions: First, we accurately identify and isolate three types of price dispersion originating from either third-degree price discrimination or peak-lead pricing. Second, we test the relative contribution of industry-elasticity and cross-price elasticity to price dispersion. Results suggest that both cross-price elasticity and industry-elasticity are crucial in determining the relationship between price dispersion and competition. Competition matters only when industry-elasticity is relatively low. And people's preferences differ to a greater extent in purchase timing and departure date than in slot, which will affect how the three types of price dispersion respond to competition.

JEL Classification: *D21, D22, D25, D43, L11, L93*

Keywords*: Competition, Price Discrimination, Price Dispersion, Price Distribution, Airlines*

---

*We are intellectually indebted to Professor Andrew Sweeting, Ginger Jin, and Chenyu Yang. We would like to thank all participants during the Summer IO Weekly Meeting at UMD for their helpful comments. We are grateful to the anonymous airline for giving me access to the data used in this study and appreciate their help in providing background information in the airline industry in China. All errors are our own.

†Georgetown University Email:yf170@georgetown.edu

§University of Maryland Email:qwang819@umd.edu

1 Introduction

Understanding the price differential among related products or among consumers of the same product has long been a pursuit of economists. What's more intriguing is what happens to such price differential after a change in the competition landscape of market. In this paper, we review the effect of competition on price dispersion [1] in the airline industry of China. Leveraging a detailed data which is also at a more granular level than those used in past studies, we are able to identify two types of price dispersion and to study the effect of competition on these two types separately. This paper has three objectives: First, quantify the degree of fare inequality in the airline industry in China. Second, describe the patterns of different types of price dispersion (discrimination-based and cost-based [2]) across different markets. Third, test the theory that predicts the relationship between competition and price dispersion.

Airline industry is a vital industry in our modern economy with its industry value as high as billions of dollars. Increasing number of travellers in China choose air-travel as their preferred transportation option. The total number of airplane passengers has tripled from 2009 to 2019 and reached 660 million in 2019. On the one hand, airfare is famous for its sophisticated pricing mechanism and it is well known that the money a passenger pays for her seat on a flight can be very different from the amount that the passenger sitting next to her pays. Understanding the source of dispersion in airfare is an important step toward unraveling the welfare implication of airfare design. On the other hand, when it comes to assessing the welfare implication of a change in the market competition condition, like a merger or divestiture, understanding the effect of competition on airfare dispersion is crucial. As Lazarev et al. (2022) argues, in traditional merger analysis, our main focus is the change in the average price after an upheaval in the market structure like M&A. Empirical

[1] As stated in Borenstein and Rose (1994), "price dispersion here refers to the variation of price charged to different passengers on the same airline and route."

[2] The concept of cost here is close to "shadow cost".

welfare implication based on average price could be misleading. After a merger of two players in an industry, different consumers may face different price changes. Even if the average price rises after a merger happens, such increase doesn't necessarily entail welfare loss due to the existence of price dispersion. The welfare gain from consumers facing a lower price may outweigh the welfare loss from consumers facing a rise in product price. Theorists have proposed several theories to make predictions of where price dispersion comes from and many empirical economists also provide evidence of the effect of competition on price dispersion. But little empirical work can be found to study the effect on different types of price dispersion due to data limitation.

Several theories have been proposed to illustrate the source of price dispersion in the airline industry. Building on Prescott (1975) and Eden (1990), Dana Jr (1999) predicts low price tickets will be provided first. After low price tickets are sold out, only then high price will be available to late-arrive travellers. This implies, in low demand state, only low price tickets will be sold. In high demand state, both low price tickets and high price tickets will be sold. In Gale and Holmes (1993), advance-purchase discount is present. Travellers with lower cost to take non-preferred flight will buy in advance and enjoy lower price. Low-cost-of-waiting travellers will be shifted to off-peak flight. Both Dana Jr (1999) and Gale and Holmes (1993) share the view that there is a pricing mechanism designed by airlines in response to different demand states and such pricing design will imply intra-firm price dispersion. Borenstein and Rose (1994) call this type of pricing "peak-load pricing" as it reflects the shadow cost of remaining seats. This will cause the price differential among passengers on the same flight or on the same flight number among different days.

An alternative theory argues that price dispersion in airfare is an embodiment of price discrimination. In Borenstein (1985) and Holmes (1989), such discrimination is close to the third degree price discrimination and consumers are discriminated against by willingness-to-pay and degree of brand loyalty. Dana Jr and Williams (2022) focus on another type of price discrimination, intertemporal price discrimination due to dynamic pricing. They stress that intertemporal price discrimination is possible in oligopoly when inventory control is possible.

When it comes to the effect of competition on price dispersion, Borenstein (1985) and Holmes (1989) predict that price discrimination may increase as a market moves from monopoly to imperfect competition. In both of their models, consumers are segmented based on their cross-price elasticity of demand among brands and this will generate more price dispersion when a market becomes more competitive. The effect of competition on price differences depends on whether discrimination is based on differences in consumers' willingness-to-pay or differences in degree of brand loyalty. Stole (2007) shows that the relationship between third degree price discrimination based price dispersion and competition is determined by the degree of heterogeneity in consumer' demand elasticity. Specifically, if consumers all have high cross-price elasticity of demand among brands, then competition would press the price toward marginal cost and thus would reduce price dispersion. If cross-price elasticity varies greatly among consumers, price dispersion will increase with competition since competition would raise prices for the former type of consumers while lower prices for the latter type of consumers.

The airline industry is a great test field for the theories mentioned above. People have found mixed empirical evidence of the relationship between competition and price dispersion in the airline industry. Borenstein and Rose (1994) find more competition is associated with higher price dispersion in the U.S. airline industry. Gerardi and Shapiro (2009), however, find the opposite relationship. They use quarterly panel data of the U.S. airline industry, in contrast to the cross sectional data in Borenstein and Rose (1994), and include both legacy and low-cost carriers, the latter of which are left out in Borenstein and Rose (1994). Dai et al. (2014) find a non-monotonic relationship between competition and price dispersion. Stavins (2001) finds that competition will stimulate price dispersion which is due to ticket restrictions. Gaggero and Piga (2011) achieve the same conclusion as Gerardi and Shapiro (2009) in the context of Irish airline industry .

There are also some papers studying the effect of competition on price distribution and the relationship between competition and airline's dynamic pricing behavior. Hernandez and Wiggins (2014) find airlines will compress their menu of fares when facing competition from low-cost carriers like Southwest. Chandra and Lederman (2018) find that competition has little impact on the top or bottom of price distribution while has a significant impact on the middle part of price distribution. Thus some price dispersion increases while others decreases. In contrast, Gerardi and Shapiro (2009) find that lower-end fares are more responsive to competition than higher-end fares. Hortaçsu et al. (2022) find that dynamic pricing will soften competition especially when booking time is close to departure date because demand becomes more inelastic when departure is coming.

The price dispersion studied in the literature is usually at route-carrier-month level. The issue with such bulky aggregation is that price dispersion from different sources are mixed together. If competition alters different types of price dispersion in different directions, then the relationship between competition and such aggregated price dispersion could go either way, which may explain why we see conflicting results in the literature. Although Borenstein and Rose (1994) and Gerardi and Shapiro (2009) try to address the issue by adding proxies of different sources as control or running regression on different subsamples, their attempt is only a makeshift and far from satisfactory in terms of disentangling one type of dispersion from another. They attribute their failure of price dispersion isolation to lacking the key information, flight departure time and ticket booking time, to distinguish one type of price dispersion from another.

In our data, we know the exact departure time for each flight and we know how many days in advance a ticket gets booked. Using this novel data, we make several contributions to the existing literature: First, we isolate intertemporal discrimination based price dispersion, which corresponds to the third degree price discrimination in theory, from demand-smoothing motivated dispersion, which corresponds to the peak-load pricing in theory. Second, we study how the increase in competition affects these two types of price dispersion.

Third, we can test the model in Stole (2007) and Holmes (1989) given di erent markets in our sample featuring di erent levels of industry-elasticity and cross-price elasticity. Overall, we demonstrate the variation in two types of price dispersion in the context of China's airline industry and answer if the relationship between competition and two types of dispersion are di erent, trying to reconcile the con icting nding in the literature. Therefore, we provide a better understanding of the e ect of increasing competition on di erent types of consumers.

Following the empirical setting in Gerardi and Shapiro (2009), we use xed e ect to control unobserved factors like airline-speci c factors that could contaminate our results and use instruments to address the endogeneity concern. In particular, in order to better accommodate the context in China, we come up with our own instrument which is the pre-scheduled capacity. The biggest di erence between pre-scheduled capacity and enplanement, which is the instrument used in Gerardi and Shapiro (2009), is that our instrument represents the expected demand while enplanement represents the realized demand which could be correlated with the price observed in the data. To pair with the theory, we devise our empirical setting by exploiting the variation in the availability of High Speed Railway (HSR) in China. Stole (2007) makes some prediction on the relationship between competition and price dispersion based on demand elasticity. Given the variation in the availability of HSR across routes and HSR's competitiveness across markets, our sample is capable of constructing samples that have di erent industry-elasticity and cross-price elasticity, which Chandra and Lederman (2018) fail to do when they build their empirical work on the theory in Stole (2007). We sort all the ight routes by ight distance and consumer composition. Our results show that industry-elasticity and the heterogeneity of consumers' cross-price elasticity together determine how competition impacts price. Competition only matters for price dispersion when the industry-elasticity is not so big that dominates the cross-price elasticity for all consumers. When there is low industry-elasticity and big heterogeneity in consumers' cross-price elasticity, more competition will lower the price dispersion. But competition has di erent impacts on price distribution for di erent types of price dispersion.

The structure of the rest of the paper will be as follows: In Section 2, we will introduce our theoretical framework and o er predictions. Section 3 contains the institutional background in China's airline industry and the details of our data. In Section 4, we discuss di erent sources from which price dispersion may derive and walk through what the literature have done. Section 5 illustrates how we isolate di erent types of price dispersion and provide the summary of our data and main variables of interest. Section 6 performs the regression analysis. Section 7 engages in resolving the endogeneity problem. We present our conclusion in Section 8.

2 Model

First, we adopt the basic theoretical framework in Holmes (1989). Two rms, m and n, provide similar but di erentiated products. There are two types of consumers. Both rms practice third degree price discrimination based on the consumer type. In the context of the airline industry, di erent types of consumers di er in their purchase behaviors that are observable to airlines, for example, the timing of their purchase, the history of searching for ights, and departure time preference. Two assumptions are made in Holmes' work. First, rm m's demand by a given type i when it sets a price p_1 and n sets p_2 is the same as rm n's demand by type i when prices are reversed, $D^i_m(p_m = p_1; p_n = p_2) = D^i_n(p_n = p_1; p_m = p_2)$. Second, a unique equilibrium to the price game is assumed to exist. In the equilibrium, both rms set the same price for a give type of consumer. Under these two assumptions, Holmes shows that the demand by a certain type of consumer for a given rm's output has an elasticity that equals to the sum of an industry-elasticity component and a cross-price elasticity component. For a given rm, say m, the elasticity of m's demand by consumer of type i is:

$$e^{\text{Firm}}_{m;i}(p) = e^{\text{Industry}}_{m;i}(p) + e^{\text{Cross Price}}_{m;i}(p)$$

The industry elasticity, e^{Industry}, measures the tendency of consumers to choose outside option if the level of prices in the market goes up. The cross-price elasticity, $e^{\text{Cross Price}}$, measures the tendency of consumers to switch sellers if one product's price changes. Holmes establishes that the equilibrium price $p_i^?$ for each type of consumer satis es the following relationship:

$$\frac{(p_i^? \quad c)}{p_i^?} = \frac{1}{e^{\text{Firm}}(p_i^?)} = \frac{1}{e^{\text{Industry}}(p_i^?) + e^{\text{Cross Price}}(p_i^?)}$$

The equation above implies that price discrimination can be based on di erences in industry-elasticity between di erent types of consumers as well as di erences in cross-price elasticity between di erent types of consumers.

Based on Holmes's framework, Stole (2007) elaborates why the relationship between competition and price dispersion is ambiguous. If di erent products are close substitution to each other for all consumer types, which implies a high cross-price elasticity for every consumer (i.e. suppose an extreme case for two types of consumers, Type 1 and Type 2, $e_1^{\text{Cross Price}}(p_1^?) \quad +1$ and $e_2^{\text{Cross Price}}(p_2^?) \quad +1$), then erce competition will press both prices that target di erent types toward marginal cost and thus price dispersion across consumer types will be negligible given a constant cost. On the contrary, if the cross-price elasticity di ers a lot across consumer segments (i.e. suppose Type 1 considers two products being close substitutes $e_1^{\text{Cross Price}}(p_1^?) \quad +1$, while type 2 exhibits strong brand loyalty $e_2^{\text{Cross Price}}(p_2^?) \quad 0$), then rms will choose near-to-cost competitive price for Type 1 consumers and charge Type 2 consumers a close-to-monopoly price. Competition in this setting will give rise to bigger price di erential than that in a monopoly market.

In the airline industry, cross-price elasticity, which measures to what extent the change in one ight's price would a ect a traveller's demand for another ight, varies across travellers of di erent types. In the literature, consumers are sorted into two types: business traveller and leisure traveller. Business travellers are usually assumed to be more loyal since they are more likely to hold the membership of a frequent yer program, thus they have higher brand loyalty and are less price-sensitive. Leisure travellers are considered being price-sensitive and they don't have strong brand preference. In the context of airline industry in China, there are also big variation in industry-elasticity, which measures to what extent the change in the price of air travel would a ect traveller's demand for other transportation options like train or car. Industry-elasticity of demand for ights correlates with the degree of competition from HSR. For a given trip from city A to city B, the more appealing HSR is, the higher industry-elasticity airlines face on the same route.

We make attempt to nest the theoretical framework into a testable empirical setting. First, we sort ight routes into 3 groups based their ight distance: below 700km, short-distance routes; 700-1500km, medium-distance routes; and above 1500km, long-distance routes. Surprising may it seem, for a trip whose distance falls into the rst group, HSR is an extremely appealing outside option for almost all travellers [3] [4]. Airplane doesn't have an obvious edge in terms of travel time on such short routes since HSR's speed can be as high as 350 km per hour. Moreover, HSR is extremely punctual at minute-level, robust to weather conditions, price-stable, close to the downtown area of a city, and has much simpler check-in and safety-check procedure (Fang et al., 2020). So ights operating on such short routes are all confronted with high industry-elasticity of demand. For ights with distance between 700 and 1500 km, the competition between HSR and aircraft is shoulder to shoulder. HSR still dominates in terms of punctuality, cheapness and boarding cost. Aircraft outpaces HSR in terms of travel time, cabin environment, and services. Industry-elasticity may vary across di erent travellers. When travel distance is above 1500 km, the drawback of long travel time on HSR outweighs HSR's other bene ts and HSR also loses its

[3]https://simple ying.com/china-high-speed-rail-impact/

[4]Wang et al. (2020) and Fang et al. (2020) both nd that HSR poses serious competition to ights, in particular to short-haul ights.

advantage in price. Flight routes over 1500km have much lower industry elasticity for all types of travellers.

As for cross-price elasticity, if you are a leisure traveller, you are potentially sensitive to price and insensitive to brand name, thus have high cross-price elasticity. But a business traveller is more likely to be loyal to certain brands to get mile-credit and she is less sensitive to price change. Therefore, we exploit the variation in the share of business travellers across routes. We classify routes into tourist routes where there are mainly leisure travellers with high cross-price elasticity, big city routes where there expects to be a su cient blend of business travellers and leisure travellers, and economic city routes where a higher share of business travellers is assumed to exist than the other two groups[5].

In short, short-distance routes and long-distance routes both have homogeneous industry-elasticity while the former's is higher and the latter's is lower. Medium-distance routes are thought to have more heterogeneity in industry-elasticity. Tourist routes have the least heterogeneity in cross-price elasticity but has the highest level of cross-price elasticity. Big city routes have the most heterogeneity in cross-price elasticity. Economic city routes are regarded as having lower cross-price elasticity than tourist routes and big city routes. According to the model introduced earlier, long big city routes have uniformly low industry-elasticity but own heterogeneous cross-price elasticity across consumer segments. In contrast, medium- and short-distance tourist routes have higher industry-elasticity and also higher cross-price elasticity. Therefore, we predict that:1. In terms of third degree price discrimination based price dispersion, big city routes have bigger price dispersion than tourist routes as competition is increasing. 2. Price dispersion in medium-short routes are more responsive to competition than long routes. 3. Peak-load pricing based price dispersion shall react to competition in a similar manner no matter what kind of route is considered. We summarize the prediction based on the model above in Table (1). Since the model above

[5]We will provide the detailed de nition of tourist routes, big city routes, and economic city routes later in the empirical part.

only discusses what happens to the third degree price discrimination based price dispersion when competition goes up, we leave the prediction related to peak-load pricing based price dispersion in question and see what the data will tell us. But the di erence between big city routes and tourist routes in the e ect of competition should be lower for the peak-load pricing based price dispersion than for the the intertemporal price discrimination based price dispersion. As we will discuss later, monitoring price dispersion can only deliver a relatively vague idea of what happens to the whole price distribution when competition conditions alter.

Several assumptions are implicitly made in the discussion above: On the supply side, rstly, rms are able to conduct third degree price discrimination but don't practice second degree price discrimination. Second, cost is constant for each rm. Third, products that are pooled together to compute price dispersion are symmetric. In reality, we do nd empirical supports for these assumptions. The complete menu of fare codes is never presented to consumers. Airlines choose which subset of fare codes are available to consumers based on characteristics of consumers' booking behavior, like days remaining before departure, channel of booking, day of the week, and so on. Such pricing practice is more close to the third degree price discrimination rather than the second degree price discrimination. [6]. As for the cost, aviation fuel cost is a major constituent of the ight operation cost. The market price for aviation fuel during our sample period is stable. As far as the symmetry assumption is concerned, it is not absurd in our empirical setting because, as shown later, our price dispersion is calculated among ights operated by the same carrier and on the same route, call these ights an airline-route-speci c ight bundle. There is no quality di erence or brand name heterogeneity within the bundle. Admittedly ights within the same bundle may depart at di erent times. Actually, the symmetric assumption we make

[6] A conventional view in the literature is that, if there are more than 20 bookings made within one day, then group buying behaviors are thought to be exist. In our data, the proportion of ight*booking-day*cabin that has more than 20 tickets booked is only 0.1% and the proportion of ight*booking-day*cabin that has more than 5 tickets sold is only 0.8%, which indicates group buying is rare in our sample. Therefore ignoring second degree price discrimination won't be a big departure from the reality.

can be interpreted as that people who buy ights in bundle t but departing at time A and people who buy ights in bundle t but departing at time B are from two di erent markets. The ights in bundle t are the same or symmetric and the price dispersion in bundle t results from di erent prices charged at di erent markets. On the demand side, we abstract away from potential strategic behavior of consumers which is discussed in some literature (Board and Skrzypacz, 2016; Li et al., 2014). So there is no self-selection in our model.

3 Institutional Background And Data

Institutional Background

The airfare in Chinese airline industry is determined by full price and discount rate. Full price varies across route, airline, and ight season. There are two ight seasons within a year, "Summer Season" usually ranges from late March to late October while "Winter Season" ranges from late October to late March in the next calendar year. For a given route, each airline will only have two chances to adjust its own full price within a calendar year, which are allowed at the start of "Summer Season" and "Winter Season". The menu of discount rate assigned to each cabin is more stable than that of full price. Each airline has its own menu of discount rate and this discount rate menu varies only at airline level and doesn't change over time nor across routes. For a given airline, each discount rate corresponds to a cabin class on the ight. There are 26 cabin classes[7] on a ight, each of which is labeled with a letter. So each discount rate or cabin class essentially is the fare code discussed in the U.S. airline industry setting. We will use discount rate, cabin class, and fare code interchangeably in this paper. Passengers under the same fare code are assumed to pay the same price.

[7]You should think of the cabin class here as "fare code" in Chandra and Lederman (2018) rather than economy/premium economy/business/ rst-class . There may exist heterogeneity in restrictions associated with di erent fare code categories, but here we make the assumption that these restriction di erences won't a ect demand. It makes sense in a way that most restrictions are about earning miles and luggage quota. Earning miles through the "Frequency Flyer Program" is much less popular than that in the western world. And most airlines in our sample are legacy airlines that provide full service for all fare code categories, which means in general boarding passengers enjoy su cient luggage quota.

Along with the list of full price, the ight schedule for Chinese airlines also changes twice a year, one at the beginning of "Summer Season" and the other at the beginning at the "Winter Season". Once entering a ight season, the full price is xed through the whole season for a given airline-route. When selling tickets over the booking horizon, airlines face capacity constraint and stochastic demand. Dynamic pricing is applied to cope with the pro t maximization problem under uncertain demand and xed inventory. On each day of the booking horizon before departure, airlines make decisions on which cabin class and how many seats in the cabin class are available to consumers. So prices can vary over the booking horizon and across cabins.

Data

The data we use in this paper is from a global distribution system (GDS) used by all airlines in China. Our sample period ranges from 2018-10-01 to 2018-10-25[8]. From the data we know bookings for each cabin class over the 45-day booking horizon before the departure of each ight. For example, for ight CA1831 departing on 2018-10-12, we can see its daily bookings for each of its 26 cabin classes from 2018-08-28 to 2018-10-12. An observation in the data is at Flight-Departure Date-Booking Date-Cabin level and it shows that for a given cabin class W on ight CA1831 departing on 2018-10-12, the tickets in cabin W that have been sold up to 2018-09-29 is 37. We also know the cabin W's discount rate, which is the same to all the ights of CA, is 0.7. We can recover the sale of each individual booking day from the cumulative sales. If the sales up to 2018-09-30 for cabin W on this ight is 45, then we know the net sales for cabin class W on 2018-09-29 is 8[9].

Another data source we use is the full prices of all the ights in the ight season. This data scrapes each Chinese domestic ight's daily lowest price and its corresponding discount

[8] The ending of our sample period is well picked so that the ight schedule and full price don't change over the sample course.

[9] As in Williams (2022), we will take the sales net of cancellation as if it is the true sales for cabin class W. This approximation is appropriate since the proportion of (ight*booking-days*cabin) that has cancellation is around 0.6% in our sample. The low cancellation rate attributes to the high cost of cancellation which can be as much as 90% of the ticket's face value (http://www.xinhuanet.com/politics/2018-05/04/c_1122784062.htm).

rate level from the largest online travel agency (OTA) Ctrip, which takes up 40% of the market share. Based on the listed price and its discount rate, we can calculate the full price for each ight which stays the same over the whole ight season.

With two elements together, discount rate and full price, we therefore determine the price of each sold ticket.

To balance the representativeness and computational burden, we limit our analysis to the top 531 busiest [10] routes in October 2018 in China which account for 65% of total domestic airplane passengers. Since we mainly compare our results with empirical work on the U.S. airline industry, we provide some basic information to illustrate two countries' airline industry. In 2018, four U.S. airports make it to the top 15 busiest airports in the world, ATL[11](1st), LAX(4th), ORD(6th), and DFW(15th). Three airports located in mainland China make it to top 15, PEK(2nd), PVG(9th), and CAN(13th). In our analysis, the route is directional. Beijing to Shanghai and Shanghai to Beijing are regarded as two di erent routes in our analysis, which is a common practice in literature (Chandra and Lederman, 2018). De ning routes in a directional manner is desirable in the context of China's airline industry. On the route from Beijing to Shanghai, Air China holds more dominant positions or has a stronger airport presence at the origin airport. On the route from Shanghai to Beijing, however, China Eastern is more dominant at the origin airport. Even though we can observe all domestic ights on these routes, we focus solely on tickets of coach classes on direct [12] ights, which is consistent with Borenstein and Rose (1994) and Gerardi and Shapiro (2009). We also drop observations whose discount rate is above 90%, a common practice in the literature which aims to avoid coding error, frequent yer awards, or employee discount.

[10] In terms of enplanement.

[11] The International Air Transport Association's (IATA) Location Identi er is a unique 3-letter code (also commonly known as IATA code) used in aviation and also in logistics to identify an airport.

[12] Direct service here means a passenger doesn't change planes during her trip. Although we don't distinguish nonstop services and other direct services due to data limitation, our talk with an employee in the airline makes we believe that passengers on these direct ights won't change during the entire trip.

A market in our analysis could be characterized by (Route, Departure Date, Slot), (Route, Departure Date, Booking Time), or (Route, Slot, Booking Time) depending on our purpose. By slot, we mean the time window within a day during which the flight takes off and will return to the details later. Booking time refers to how many days in advance a ticket is booked. To make our results comparable with previous literature, we focus on domestic, direct non-stop, coach-class airline tickets.

4 Sources Of Price Dispersion And Past Studies On Price Dispersion

Sources of Price Dispersion

There are several sources of price dispersion mentioned in literature, including intertemporal price discrimination, systematic peak-load pricing and stochastic peak-load pricing. [13] The first source comes from the fact that airlines offer different prices if consumers make a reservation with different days before the departure. The second and the third source are related to the airline's motivation to smooth demand across flights in response to expected demand flow and unexpected demand flow respectively. In order to cover the huge sunk cost of executing a flight, price is used as a device to allocate demand across flights[14] to make sure each flight as full as possible. The previous studies can not really map their empirical setting to the price dispersion theory due to the data limitation. With data at more aggregate level (i.e. usually Route-Carrier-Quarter level data from DB1B[15] in the U.S. case), Gerardi and Shapiro (2009) and Borenstein and Rose (1994) are incapable of isolating different types of price dispersion coming from different sources. The price dispersion they measure is actually a mix of different types of price dispersion. If different types of dispersion don't react to changes in competition in a uniform manner, then the relationship between competition and price dispersion at a more aggregated level is ambiguous. In

[13] As defined in Borenstein and Rose (1994), `systematic' peak-load pricing reflects variations in the expected shadow costs of capacity at the time a flight is scheduled. `stochastic' peak-load pricing refers to demand uncertainty for individual flights that is resolved only after equipment scheduling decisions are made.

[14] Consumers with lower ticket price, lower waiting cost or weaker flight preference are usually the target to be shifted.

[15] Details about this data can be found in Borenstein and Rose (1994) and Gerardi and Shapiro (2009).

contrast, our data has a higher frequency, daily level rather than quarter level, and consists of more detailed information on each flight and each booking. These advantages motivate us to do a better job in disentangling different sources of flight fare dispersion. Before our decomposition, it is necessary to walk through past studies on price dispersion in the airline industry.

Comparison With Literature

Borenstein and Rose (1994) use cross sectional fare variation in DB1B data and find price dispersion increases in competition but decreases in flight frequency on a route. They interpret this result as evidence for lower-end price being more responsive to competition than higher-end price. Heterogeneity in cross-price elasticity (i.e. brand loyalty) plays a major role in determining price dispersion. Only legacy carriers are considered here. Like us, they focus on domestic, direct non-stop, coach-class airline tickets.

Gerardi and Shapiro (2009) also focus on domestic, direct non-stop, coach-class tickets and also use DB1B data which is on a quarterly basis. With a long panel data structure, they reach the opposite conclusion. They find a negative relationship between competition and price dispersion. They include both legacy carriers, Low-Cost carriers, and regional carriers. Their sample ranges over a longer period during which there are many entry and exit events happening over the course of their sample. But these differences between their samples are not the reason for the disparity in their conclusions. Gerardi and Shapiro (2009) argue that the conflicting results result from the troublesome identification strategy in Borenstein and Rose (1994).

While Gerardi and Shapiro (2009) have great time-series variation in competition which we can't match, they miss information on the exact time of departure. Therefore, peak-load pricing and demand pricing patterns can not be detected directly. In Borenstein and Rose (1994) they make attempt to control for peak-load pricing by some proxies, while Gerardi and Shapiro (2009) take another route. They distinguish between big city sample and leisure sample [16]. The latter type of route mainly consist of passengers with high elasticities of demand and low reservation prices, called leisure travellers. Big city routes, however, feature a mix of leisure travellers and business travellers who have high reservation prices. They argue that if price dispersion mainly roots from price discrimination, then big city routes should be more affected by changes in competition than a route with a more homogeneous customer base. But peak-load pricing and stochastic demand implementation should not be affected by route type. We don't find their argument very convincing. People's preference for departure time can also be dependent on route type. For example, travellers whose destination is a tourist city are more likely to travel during the holiday season.

We will introduce Gerardi and Shapiro (2009)'s regression specification in detail here since it is their result that we intend to compare directly. First, they compare the median of price dispersion by (Big City, Leisure)*(Monopoly, Competitive) [17]. Then the panel regression of price dispersion on competition is performed where each observation is carrier i, route j, and quarter t.

$$G^{lodd}_{ijt} = \ln \frac{G_{ij}}{1 \quad G_{ij}} = \quad + \quad Competition_{jt} + \quad X_{it} + \quad {}_{t} + v_{ij} + \quad {}_{ijt}$$

where G^{lodd}_{ijt} is Gini log-odds ratio. X_{it} is an indicator of whether airline i is in bankruptcy at time t. $Competition_{jt}$ have 4 measures: First, the negative of the logarithm of Herfindahl-Hirschman Index (HHI). Second, the logarithm of total number of carriers. Third, number of legacy and low-cost carriers respectively. Fourth, dummy variables indicating the market structure being monopoly or oligopoly.

They also focus on the effect of competition on price distribution. Changes in price dispersion can be resulted from many things. For example, either the upper portion rises to a greater extent than the lower portion or the lower portion falls by a greater extent than the upper portion. Changes in the distribution of price are more informative.

[16] Their leisure sample is conceptually similar to our tourist sample.

[17] They call markets with more than one carriers "competitive" markets while as shown later we call those with more than two carriers "competitive" markets.

$$\ln P(k)_{ijt} = \quad + \quad Competition_{jt} + \quad X_{it} + \quad {}_{t} + v_{ij} + \quad {}_{ijt}$$

where k is either the 10th or 90th percentile. All the regressions above are running by subsample: big city routes and leisure routes. In order to address the concern of endogeneity in competition, they adopt several instruments to conduct Two Stage Least-Squares (2SLS) regression. They try the instruments used in Borenstein and Rose (1994) rst and argue that they may not be valid. Then they propose their own instrument: the logarithm of total enplaned passengers for a given route and period, the arithmetic mean of the metropolitan population of end-point cities, and the geometric mean of the metropolitan population of end-point cities.

Chandra and Lederman (2018) studies the Canadian airline industry. They have only one carrier, Air Canada, in their sample. The impact of competition on the overall amount of fare distribution, not fare dispersion, on a route is their focus.

$$\log P^{i}_{rt} = \; {}_{0} + \; {}^{i}_{1} Competition_{rt} + \; {}_{r} + \; {}_{t} + \; {}_{rt}$$

where r indexes route and t indexes month. i denotes di erent types of fares on a given route. For example, the average coach or average business class fare, or else speci c percentiles of the overall fare distribution. An observation is a route-month combination. Standard errors are clustered at the route level. They express prices in logarithm to measure the proportional e ect of competition on various fare measures. They use fare code to approximate the empirical distribution of fares. Chandra and Lederman (2018) mention that two reasons would give rise to the situation where there is variation in fares across passengers within a fare code. One is the airline changes prices of fare code over the booking horizon. The other one is that airlines may set di erent prices for the same fare code across di erent ights on a route. Since neither situation happens in China, we do not su er the same potential measurement error as Chandra and Lederman (2018) in constructing price distribution when using fare code to approximate. Competition is measured by: (i) number of carriers other than Air Canada; (ii) indicators for whether the market is a duopoly or competitive; [18] (iii) the negative of the logarithm of the HHI in the route-month.

5 Price Dispersion Decomposition

Like we said in the previous section, past studies have made unsatisfactory attempt to isolate price dispersion from different sources due to data limitation. Borenstein and Rose (1994) take different sources of price dispersion into account by adding proxies of each source. Gerardi and Shapiro (2009) isolate the price discrimination source and the non price discrimination source by comparing results from different types of route. Chandra and Lederman (2018) propose a theoretical framework to predict the price differentials based on Stole (2007) but conduct an empirical setting loosely connected with their model. What we do in the following has the flavor of decomposition which is in a conceptual sense but not in a mathematical sense.

Variable Construction

Following the literature (Borenstein and Rose, 1994; Gerardi and Shapiro, 2009), we use Gini coefficient to measure price dispersion. A Gini coefficient describes the inequality across the entire range of airfare paid. The larger the Gini coefficient is, the bigger price inequality a sample has. For example, the median of Gini coefficient in China's airline industry is 0.143, which implies an expected absolute fare difference of 28.6 percent of mean fare for two passengers selected at random on a given route, carrier, and departure date.

Even though we do not have as many exit and entry events as Gerardi and Shapiro (2009), we have a sharp edge over past studies in terms of measuring price dispersion at more granular levels, which enable us to disentangle different theories in our empirical setting. Specifically, Gerardi and Shapiro (2009) don't know any of the three information

[18] Their definition of "competitive" market is the same as ours.

of each flight: departure date, departure time, and each ticket's booking time, all of which we can observe in our data.

First, following Gerardi and Shapiro (2009), we measure price dispersion using Gini coefficient for a group of tickets that share the same carrier, route, and departure date and we call it G^{GS}. Second, we calculate Gini coefficient for a group of tickets that share the same carrier, route, slot [19], and departure date. By the way of construction, we call this Gini coefficient "Intertemporal Price Discrimination Based Gini" (IPD Gini G^{IPD}) because it essentially captures the price dispersion over the booking horizon prior to departure. In Williams (2022) and Hortaçsu et al. (2022), they find significant variation in willingness-to-pay across days from departure and show that demand becomes more inelastic as departure day is getting closer not only on the monopoly route but also on the oligopoly route. Our second measure of price dispersion will specifically capture any pricing behavior that exploits such variation in demand. Third, we calculate another two types of Gini coefficient which are constructed at (Carrier, Route, Booking Time, Slot) level and (Carrier, Route, Booking Time, Departure Date) level respectively. Both of the two types of Gini coefficient try to measure the price dispersion which comes from peak-load pricing theory. We call these two Gini coefficients G^{slot} and G^{date}. The latter one measures how dispersed airfare is designed at a given booking time by one carrier in order to allocate demand across departure dates for flights on the same route and taking off in the same slot. For example, suppose two flights are operated by the same airline on the same route and take off both in the morning slot, the only difference is that one departs on December 24 while the other departs on December 25. Usually flights on December 24 are on higher demand than flights on December 25 since people try to get home before Christmas Eve and not many people fancy the idea of traveling on December 25 conditional on both tickets being affordable. In order to smooth the demand between these two flights, at a given time in the booking horizon, say December 10, the carrier may charge higher price for the December 24 flight

[19] The departure time of the flight, like 10:30 am.

and charge lower price for the December 25 flight. For those who are on a tight budget, they probably choose the flight departing on Christmas due to a lower price. The former one G^{slot} measures how dispersed airfare is designed by one carrier at a given booking time in order to allocate demand across slots for flights on the same route and taking off on the same day. Some people are willing to catch early flight if choosing the early flight means saving money. In Figure (1) we pick one route PEK-SHA (Beijing to Shanghai) from our sample and illustrate G^{IPD}, G^{slot}, and G^{date}.

Specifically, slot is a categorical variable which indicates the time block of a flight's departure time or takeoff time. Slot can be early morning (before 7am), morning (7am-1pm), afternoon (1pm-7pm), or night (after 7pm). Booking time specifies how many days in advance a ticket is booked. We split the booking horizon into 5 groups based on their gap to departure date: within 5 days, 5-10 days, 10-15 days, 15-20 days, 20 days and beyond.

Subsamples Of Interest

As we mention in Section 2, the airline industry in China provides a rare test field for price dispersion theories. One feature is that we have great variation in industry-elasticity across flight routes. HSR is a very competitive alternative to aircraft for short-medium distance travel. So based on flight distance we divide our sample into 3 groups: short-distance routes, medium-distance routes, and long-distance routes. Travellers on routes in the first group have high industry-elasticity of air-trip demand while those on routes in the third group have relative lower industry-elasticity due to the material advantage of aircraft in terms of travel time. The second group have the biggest heterogeneity in industry-elasticity because both travelling means can be appealing to some people while unattractive to others at the same time.

Also we split the routes based on the degree of heterogeneity among travellers or the likelihood of a traveller being business traveller. Following Gerardi and Shapiro (2009), we define a route to be tourist route or leisure route if one of the route's end point ranks top 20 in terms of tourism revenue-GDP ratio in 2019[20]. We de ne a route to be big city route if both of the route's end points are ranked top 20 in terms of population in 2019. The idea is that we presume tickets sold on leisure routes are mainly bought by leisure travellers who have higher cross-price elasticity of demand and lower reservation price. On big city routes, travellers are presumed to be a mix of business travellers and leisure travellers. Besides the two types of routes (tourist and big city) in literature, we also split the sample based on the GDP of end-point city of a route. If a route's origin and destination both rank top 20 in terms of GDP in 2019, we label this route as economic city routes. It is worthy of pointing out that the rank of GDP is not positively correlated with the rank of population. So the economic city routes are distinct from big city routes and economic city routes have larger share of business travellers than the other two groups. In short, We are implicitly assuming that the shares of business travellers in tourist routes, big city routes, and economic city routes are in ascending order while big city routes have the greatest heterogeneity in demand in terms of price sensitivity and preference. We mark the big cities, tourist cities, and economic cities on a map (Figure(3) in Appendix C) so you can have a better understanding of the geographical distribution of these markets.

Summary Of Samples And Variables

First, we want to know the demand dynamics for ights at di erent times. The load factor is calculated for each ight by dividing the number of sold seats when departing by the seating capacity. Then we track the load factor across departure dates, across slots, and over the booking horizon. As shown in Figure (2), demand uctuates a lot over the departure dates. To what extent the aircraft is lled is also changing dramatically over the booking horizon. In contrast, ights taking o at di erent times during the day don't show a big variation in their load factor. These patterns of demand will greatly assist us in

[20] You may notice we use the rank of tourism revenue-GDP ratio, population, and GDP in 2019 instead of in 2018, which is due to data limitation. In our favor, however, it is very unlikely that these ranks change signi cantly within 2 years.

understanding price dispersion.

Second, we focus on the patterns of price dispersion in di erent markets:

From Table (2), we can see that monopoly markets, duopoly markets and competitive[21] markets mainly di er in the size of enplanement. The median of enplanement in competitive market is as ve times many as that in the monopoly market and is more than double size of that in the duopoly market.

Table (4) shows that on average price dispersion across departure dates is the biggest. Price dispersion from intertemporal price discrimination, G^{IPD}, comes second but is lower than the price dispersion measured in the same way as Gerardi and Shapiro (2009),G^{GS}. Notably, G^{slot} is much smaller than other price dispersion, which may be explained by the almost-uniform demand distribution across slots in Figure (2). In comparison, the median of Gini coe cient in China's airline industry, 0.143, is lower than 0.18 in Borenstein and Rose (1994) and 0.22 in Gerardi and Shapiro (2009). Our measure is at daily frequency while they calculate Gini at quarterly frequency.

Table (6) shows that longer routes generally tend to have higher price dispersion. If we focus on long routes (distance greater than 1500km) where there are little competition from outside option HSR, we can see di erent relationships between competition and different price dispersion. G^{GS} decreases in competition[22], which is consistent with existing evidence in U.S. airline industry. As for G^{IPD} which embodies intertemporal price discrimination, G^{slot} which is supposed to capture the motivation of smoothing demand across slots, and G^{date} which is supposed to capture the motivation of smoothing demand across dates, however, all display a non-linear relationship with duopoly markets displaying the highest price dispersion.

If we break the whole sample into big city routes, tourist routes, and economic city routes, the negative relationship between competition and G^{GS} still exists if we compare

[21] There are more than two airlines.

[22] We ignore the long-monopoly routes here since the sample size is too small to make the statistic representative.

duopoly markets with competitive markets. If intertemporal price discrimination is the main determinant of price dispersion, big city routes which have a more heterogeneous consumer base should have higher G^{IPD}, a type of dispersion solely based on price discrimination, than tourist routes which are thought to have a more homogeneous consumer base. G^{IPD} in big city routes also are expected to be more a ected by competition than that in tourist routes. In Table (7), we don't see G^{IPD} is higher in big city routes than in tourist routes. Peak-load pricing theory predicts that price dispersion is a re ection of price's response to demand and won't be a ected by route type. On the contrary, Table (7) shows that there exist non-trivial di erences in both G^{slot} and G^{date} between big city routes and tourist routes. Speci cally, G^{slot} is larger in big city routes and economic city routes that that in tourist routes. G^{date}, however, is higher in tourist routes than that in big city routes and economic city routes. The descriptive pattern is consistent with our intuition that for routes with more tourist, departure date is the main driver of the demand while demand of business travellers are more sensitive to the take-o time.

6 Regression

Empirical Setting

As argued in Gerardi and Shapiro (2009) and Chandra and Lederman (2018), although Gini coe cient is a useful and normalized measure of dispersion, a change in Gini could be an outcome caused by di erent changes in price distribution. Gini alone doesn't reveal much information about the price distribution dynamics. Thus, apart from Gini coe cient, we construct 10th and 90th percentile of airfare for a given group of tickets and use them as left-hand side variables. For each type of price dispersion, we have 4 regression equations to investigate the e ect of competition. Two focus on Gini coe cient and two focus on price at certain percentile. The reason we have two regressions for Gini coe cient or price level separately is that the two measures of competition vary at di erent aggregation level and thus have di erent subscripts.

$$\ln G^{IPD}_{ijst} = {}_0 + {}_1C_{1;jst} + {}_{ij} + {}_t + v_s + u_{ijst} \tag{1}$$

$$\ln G^{IPD}_{ijst} = {}_0 + {}_1C_{2;js} + {}_{ij} + {}_t + v_s + u_{ijst} \tag{2}$$

$$\ln P(k)_{ijst} = {}_0 + {}_1C_{1;jst} + {}_{ij} + {}_t + v_s + u_{ijst} \tag{3}$$

$$\ln P(k)_{ijst} = {}_0 + {}_1C_{2;js} + {}_{ij} + {}_t + v_s + u_{ijst} \tag{4}$$

$$\ln G^{slot}_{ijbt} = {}_0 + {}_1C_{1;jt} + {}_{ij} + {}_t + v_b + u_{ijbt} \tag{5}$$

$$\ln G^{slot}_{ijbt} = {}_0 + {}_1C_{2;j} + {}_{ij} + {}_t + v_b + u_{ijbt} \tag{6}$$

$$\ln P(k)_{ijbt} = {}_0 + {}_1C_{1;jt} + {}_{ij} + {}_t + v_b + u_{ijbt} \tag{7}$$

$$\ln P(k)_{ijbt} = {}_0 + {}_1C_{2;j} + {}_{ij} + {}_t + v_b + u_{ijbt} \tag{8}$$

$$\ln G^{date}_{ijbs} = {}_0 + {}_1C_{1;js} + {}_{ij} + {}_b + v_s + u_{ijbs} \tag{9}$$

$$\ln G^{date}_{ijbs} = {}_0 + {}_1C_{2;js} + {}_{ij} + {}_b + v_s + u_{ijbs} \tag{10}$$

$$\ln P(k)_{ijbs} = {}_0 + {}_1C_{1;js} + {}_{ij} + {}_b + v_s + u_{ijbs} \tag{11}$$

$$\ln P(k)_{ijbs} = {}_0 + {}_1C_{2;js} + {}_{ij} + {}_b + v_s + u_{ijbs} \tag{12}$$

where i denotes the carrier, j denotes the route, s denotes the slot, b denotes the booking time, and t denotes the departure date. k stands for 10th or 90th percentile price of coach class. $\ln G$ stands for the logarithm of Gini coe cient, unlike Gerardi and Shapiro (2009) we don't use the odd ratio of Gini here because using Gini coe cient is simpler to interpret. The full sample covers ights on the most busiest 531 routes from 2018-10-01 to 2018-10-

25. When it comes to the regression with G^{slot} and G^{date}, we drop observations whose $(i; j; b)$ a liation only has one slot or one date.

There are two ways used to measure competition: One is HHI which is based on the real-time market share of sold tickets [23]; The other is the logarithm of number of total carriers. C_1 refers to the rst measure of competition, $\log(HHI)$. C_2 refers to the second measure of competition. We use di erent notation here because their variation exists at di erent levels. In comparison with the logarithm of number of carriers, $\log(HHI)$ may be more close to the competition concept in our economists' mind. A larger number of carriers doesn't necessarily guarantee a more competitive market than a market with slightly less number of carriers. A toy example is su cient to illustrate the idea: market A has two carriers which split the passengers equally. market B has three carriers but one is dominant, accounting for 90% of the market share, while the other two are fringe players, each with 5% market share. The HHI for market A is 0.5 while the HHI for market B is 0.86. In this sense, market A is more competitive or less concentrated than market B even thought market B has more carriers.

The reason we calculate the competition at Route-Departure-Slot level is that a sta working at the airline tells us that when they set prices, the competitor in their opinion are those ights whose slot time is close to theirs. So it is necessary and makes more sense to control the slot.

Note that we calculate HHI at (Route, Date) level not at (Route, Week) level. In the U.S. or Canadian setting, measuring competition at such detailed level may looks unnecessary. Daily market share of passengers do re ect real-time market conditions but information at this granular level won't be factored into U.S. airline's pricing decision since each airline only knows its own real-time booking information. But in China, all airlines have access to real-time sales and price information of other airlines. From our talk with the sta in the

[23] In regression, we actually use $\log(HHI)$ to make the interpretation of estimated coe cient straightforward. If the estimated coe cient of $\log(HHI)$ is positive, then it means competition will raise price dispersion. Gerardi and Shapiro (2009) adopt a similar practice.

airline pricing department, they do take other airlines' real-time sales into account when setting their own price. Therefore we have time-series variation in HHI to leverage here. But the number of carriers stays the same across departure dates. The way we add xed e ect terms is in the spirit of the setting in Gerardi and Shapiro (2009). We cluster the standard error at the route level because we believe that, as argued in Goolsbee and Syverson (2008), both serial correlation and correlation between the pricing decisions of multiple carriers on the same route can be present.

Preliminary Results

First, we replicate the same empirical analysis as in Gerardi and Shapiro (2009) except for our analysis is at daily level. Results are presented from Table (12) to Table (15). A detailed discussion of these results can be found in Appendix A. Then, in the next section, we point out the potential endogeneity issue in our analysis and explain in detail why the instruments in Gerardi and Shapiro (2009) don't work in our case. Finally, we o er a novel instrument and discuss the results from 2SLS regression.

7 Endogeneity And Instrument Variable

Literature has argued that there can be potential endogeneity issue with OLS (Ordinary Least-Squares) regression results. A market with higher price dispersion may be more attractive to airlines. So estimated coe cients from OLS may su er positive bias. Besides, the variation in the number of carriers [24] is constant over our sample period, so regressions only exploiting cross-sectional variation is susceptible to the concern that markets with more carriers may have some unobserved features that lead to a higher/lower dispersion consistently. We address this issue by using xed e ect controls and instrument variables (IV).

In Gerardi and Shapiro (2009), they rst argue that the instrument used in Borenstein and Rose (1994), the logarithm of ight distance, is troublesome in a sense that ight dis-

[24] Another measure of competition $\log(HHI)$, however, has time-series variation.

tance are positively correlated with the capacity of flights which is omitted in the original regression. And the capacity of flights is positively correlated with price dispersion. The bigger a plane is, the more consumers it will serve, thus more heterogeneity in consumers and more price dispersion. Gerardi and Shapiro (2009) offer another choice of instrument as a solution, the logarithm of enplanement. In the following, we will give it try first and describe the results which are quite counter-intuitive. Second, we argue that the instrument they use, the logarithm of enplanement, doesn't work in our setting. It fails the exclusive restriction criterion of being a valid instrument. We will offer our explanation of why enplanement doesn't work as a valid IV. Third, we will propose our own IV, the pre-scheduled capacity at the origin airport. After demonstrating our IV passes the relevance and exclusion restriction criterion, we will document the regression results from 2SLS (Two-Stage Least-Squares) regression based on our IV[25].

Failure Of The Instrument In Gerardi and Shapiro (2009)

First, we use the logarithm of enplanement to instrument the measure of competition in the regression of G^{IPD}. From Table (17) to Table (19), the results of 2SLS regression with the IV in Gerardi and Shapiro (2009) show that:

In many cases, the Wald F statistic is less than 10, which is a sign of weak instrument. For G^{IPD}, competition has negative effect in medium- and long- distance markets while for G^{slot} only on meidum-distance routes competition suppresses price dispersion. The estimated coefficients from big city sample is bigger and more significant that those from tourist routes sample.

If we take the logarithm of 10th and 90th percentile of prices under different market definitions as dependent variables and see the effect of competition on prices at different percentiles. Surprisingly, we find that both high-end prices and low-end prices are higher in markets with more competition and low-end prices increase more than high-end prices.

[25] For the sake of comparison, we also provide the 2SLS regressions results with logarithm of enplanement as IV in Appendix B.

Note that the sign of some of coefficients flips after switching from OLS to the 2SLS with enplanement being IV. We are cautious about interpreting this results.

Reason Of Failure And Solution

Given the suspicious results above, we suspect that enplanement is not a valid IV in our setting. So we test if it satisfies the exclusion restriction criterion. Although we can not calculate the covariance $Cov(ENP; u)$ [26] directly, we will provide some indirect evidence implying that $Cov(ENP; u)$ is not zero and positive. After regressing price dispersion on competition, we save the residuals and regress the residuals on enplanement. The coefficient of enplanement is significantly positive. We think the reason for this positive correlation could be that higher enplanement indicates higher realized demand or a boom during a business cycle. In a boom phase of business cycle, firms will elevate their price in response to strong demand. Omitting demand state indicator will lead to the invalidity of using enplanement as IV. So the realized demand which is omitted in the regression correlates with our dependent variable (i.e. 10 percentile of price) and the instrument at the same time, which violates the exclusion restriction criterion.

As a solution, we propose the pre-scheduled capacity[27] of the origin airport as our IV. In our data, we know the number of available seats on each aircraft. Specifically, to instrument $C_{1;jst}$, we calculate the sum of the pre-scheduled capacity of all airplanes in group (route j, slot s, departure date t). For example, for the HHI measure of the group (route=CAN-TNA, slot =3 , departure date = 2018-10-08), we calculate the sum of the capacity of all airplanes in group (route=CAN-TNA, slot =3 , departure date = 2018-10-08), the capacity is determined at the beginning of Summer Season, the end of March in 2018. On the relevance side, the sum of capacity indicates how many flights and how much traffic an airport can handle in the market defined by (route j, slot s, departure date t). A bigger

[26] u is the error term in $G = \beta_0 + \beta_1 Competition + \beta_2 X + u$. ENP is the enplanement.

[27] 'Pre-scheduled' means the capacity is determined at the beginning of a flight season. In our case, the flight schedule for our sample period, 2018-10-01 to 2018-10-25, is determined at the beginning of Summer Season which is the end of March.

capacity of the airport reflects its capability of coping with larger traffic (i.e. safety check, check-in etc.) and of coordinating more take-offs. And calculating the sum of the capacity of all airplanes at such granular level is necessary. Even at the same airport, different times during a day feature different levels of traffic and take-offs to handle. The more capable of handling traffic and take-off an airport is, the more carriers are likely to operate flights at the airport. On the exclusion restriction side, the flight schedule and aircraft arrangement is predetermined at the beginning of the flight season [28], which is way ahead of our sample period. The sum of capacity reflects an expectation of total demand at the airport. Price consists of an expectation term and a stochastic term. The former is based on flight distance and the expectation of demand. The latter, however, is a response to the unexpected part of the realized demand. Any price change in response to the unexpected part is independent of the predetermined flight capacity. Therefore the price is independent of our instrument conditional on flight distance and the expectation of demand, both of which are controlled by our fixed effect terms. From the data, we see that the scheduled capacity of the origin airport strongly correlates with competition. If we regress the logarithm of 10th or 90th percentile price on our instrument directly, the coefficient is small and insignificant. Therefore, we consider our novel instrument as a valid instrument for competition. In the next part, we will present the 2SLS results after instrumenting the opposite of the logarithm of HHI and the logarithm of the number of total carriers with this novel instrument.

Results From 2SLS Regression

In Table (8) we find that G^{IPD} is significantly bigger when the market is less concentrated on medium-distance routes. The positive effect of competition on price dispersion is bigger on tourist routes than that on big city routes. We group flights that share the same route, airline, departure date,and slot together. Then we investigate the effect of competition on different portions of price distribution within one group. Prices in the lower portion usually

[28]Cases like a temporary change of aircraft are very rare.

come from tickets booked in the very early stage of the booking horizon while prices in the upper portion usually come from tickets whose booking time is very close to the departure day. When the number of total carriers increases or the market is less concentrated, both the 90th percentile price and the 10th percentile price are lower and the low-end price drops by a bigger margin. The di erence in the extent to which prices on both ends drop explains the resulting bigger price dispersion. In other words, more airlines entering the market will bene t the early purchase even more than the late purchase, which is consistent with Hortaçsu et al. (2022). Hortaçsu et al. (2022) argue that high-end price over the booking horizon should be less a ected by an increase in the number of competitors because dynamic pricing practice in the airline industry will soften competition when departure day is getting closer and demand is more inelastic. The results by route type show that economic city routes tend to experience larger price reduction that the other two types. In comparison with big city routes, both the price dispersion and price level respond to the change in competition more on tourist routes, which is inconsistent with the prediction in Gerardi and Shapiro (2009).

In Table (9) only G^{slot} on long-distance economic city routes responds to the level of competition in the market and G^{slot} tends to be lower when the number of carriers is higher. In all the other scenarios we don't nd any signi cant impact of competition on price dispersion, which is not surprising. Recall that in Table (4) G^{slot} is much smaller than the other two Gini coe cients and we nd the demand barely uctuates along the slot dimension. We group ights with the same route, airline, departure date, and booking time together, thus the price distribution within one group re ects the distribution of prices across di erent slots. As the competition intensi es, the low-end (i.e. price of the non-popular departure time slot) prices increase on medium-distance routes while decrease on long-distance routes. The high-end (i.e. price of the popular departure time slot) prices, on the other hand, only climb when there are more carriers on the medium-distance routes. Overall, we nd little evidence that price dispersion across slots responds to market structure. But the prices of flights with popular slots are higher on medium routes and the prices of flights with unpopular slots are lower after the market is less concentrated. With more airlines serving the routes, people taking slots like late night or early morning enjoy more discount while people sticking to slots like late morning or afternoon will suffer higher mark-up. The results in (9) reveal that G^{slot} is robust to the competition in the market.

Table (10) shows that when the market is more competitive, G^{date} on medium-distance big city routes and on long-distance economic city routes is larger. We group flights with the same route, airline, slot, and booking time together, thus the price distribution within one group reflects the distribution of prices across different dates. The low-end price is more often found on flights departing on a non-popular day, like December 25th. The high-end price is more often found on flights departing on a popular day, like the last day of a holiday. The effect of competition on the price distribution over dates is different across portions. The lower portion are pushed left even more than the upper portion, which leads to a wider price range.

In comparison with the results from OLS, first of all, the sign of estimated coefficients in the price level regressions stays the same. Second, the 2SLS estimated coefficient of competition on high- and low-end price is smaller than that of OLS estimates in most cases, which implies a positive bias of the OLS estimates. Third, we don't find 2SLS estimates are exceedingly different from OLS estimates in terms of magnitude. The comparison gives us more confidence in our 2SLS results. But the first point doesn't hold in the regression of Gini coefficient on competition. In particular, the estimates from 2SLS are usually higher than those from OLS, but magnitude-wise the gap is not big enough to make the estimates invalid and most of the OLS estimates are insignificantly. It is likely that there is a omitted factor that is positively related to competition but negatively related to price dispersion. Such candidate can be from the demand side. For example, if the market concentration is usually correlated with the attractiveness of market. Markets with more attractiveness may have less heterogeneity in cross-price elasticity among consumers and thus less price dispersion.

Overall, a change in price dispersion is a by-product of the change in price distribution and our estimates on price distribution are reliable based on the three observations above, thus we are con dent in our 2SLS estimates.

Summary of Results

In summary, the e ect of competition on di erent types of price dispersion do vary a lot. For intertemporal discriminatory price dispersion, with more competitors or lower concentration, we nd a rise in price dispersion in the sample of medium- distance routes. A one-standard-deviation reduction in HHI from its mean (a 40% decrease) would pull up the intertemporal price dispersion by 10%, that means the expected price di erence of buying two coach-class tickets with di erent time left to the departure is 2.76% of the mean fare. For peak-load pricing motivated price dispersion, G^{slot} is less sensitive to competition than G^{date}, partly due to that there is not much uneven demand across slots to be smoothed. For price di erential across slots, the negative e ect of competition is only found on the long-distance economic city routes. For price di erential across dates, the reduction in the market concentration would lead to wider price gap between ights on popular dates and ights on non-popular dates, all else equal. A one-standard-deviation reduction in HHI from its mean (a 40% decrease) would pull up the price dispersion across dates by 24%, which means the expected price di erence of buying two coach-class tickets for ights departing on di erent dates is 8.78% of the mean fare. Our results suggest that both industry-elasticity and heterogeneity in consumers' cross-price elasticity play an important role in shaping the relationship between competition and price dispersion.

The di erence between the estimates from di erent types of routes (di erent levels of heterogeneity in the consumer cross-price elasticity) is a straightforward demonstration of the prediction that the heterogeneity in the cross-price elasticity among consumers does impact how competition a ects price dispersion. Speci cally, consumers exhibit di erent levels of preference dispersion toward di erent product features. Our nding of the positive relationship between G^{IPD} G^{slot} and competition re ects that there is great variation in consumers' preference of the timing of purchasing ight tickets and preference of departure date, even on tourist routes, which contradicts with the assumption in the literature that tourist routes have a group of customers with homogeneous preference. As for the take-o time of a ight, morning or evening, on general aircraft passengers in China don't have as much variation in their preference of take-o time as in the preference of purchase time and departure date.

When it comes to the industry-elasticity, the previous studies either understate the role of industry-elasticity or fail to test the e ect of industry-elasticity directly. As far as we know, we are the rst to demonstrate that competition has a material impact on price dispersion only on routes that are long enough. In other words, when the outside option is appealing to everyone in the market, competition between producers within the market is a trivial factor to the price dispersion. In Stole's model, this corresponds to the situation where $e^{Industry} >> e^{Cross\ Price}$ 8i, then

$$\frac{(p_i^? \quad c)}{p_i^?} = \frac{1}{e^{Firm}(p_i^?)} = \frac{1}{e^{Industry}(p_i^?) + e^{Cross\ Price}(p_i^?)} \quad \frac{1}{e^{Industry}(p_i^?)}$$

On the contrary, when $e^{industry}$ doesn't dominate $e^{Cross\ Price}$, greater heterogeneity in $e^{Cross\ Price}(p_i^?)$ is indeed a necessary condition for competition to be in uential on price dispersion.

Competition impacts di erent portions of the price distribution di erently. The price distributions over booking horizon and across dates move leftwards. Competition brings a bigger markdown to the low-end prices (i.e. early purchase or ight on a non popular date) than to the high-end prices (i.e. late purchase or ight on a popular date). So the whole price distribution moves leftwards unevenly with the lower portion being stretched more to the left, which is intuitive given the well-documented fact that demand near departure date and demand on popular dates is more inelastic. If the demand is big enough relative to the supply, the inelasticity nature of demand would soften the competition e ect. But we don't see a clear direction in which the price distribution across slots moves.

Lastly, we notice that price dispersion reacts more actively to the market concentration HHI than to the number of total carriers, which is consistent with our earlier argument that HHI is a better measurement of competition than the number of total carriers.

8 Conclusion

In this paper, we use a detailed data to revisit the e ect of competition on price dispersion. We make an innovation that price dispersion from di erent sources are isolated. The magnitude of both intertemporal discriminatory price dispersion and peak-load pricing based price dispersion is signi cant in Chinese airline industry. The expected absolute fare di erence between two passengers on the same airline, route, departure date and slot can be as high as 78% of the mean fare. For passengers on the same airline, route, booking time, and departure date, such expected absolute fare di erence on average is 18% of its mean fare. And we nd the price dispersion tends to increase in ight distance, which implies price dispersion decreases in industry-elasticity.

In order to identify the e ect of changes in competition on di erent types of price dispersion and price distribution, methodologically, we use xed-e ect and instruments in our regressions. To resolve the potential endogeneity, the instrument we propose is the sum of predetermined (10 months in advance) ight capacity at the origin airport. It turns out that our IV works well for most cases. We nd that competition will elevate the intertemporal price dispersion and across-date price dispersion only when the industry-elasticity is reasonably low.

What's more important is that we provide empirical counterpart to the theoretical work. There are three main takeaways: First, we nd empirical support for the model in Stole (2007). Especially, competition only matters for price dispersion when the industry-elasticity doesn't dominate the cross-price elasticity. So it's necessary to control for industry-elasticity when studying the e ect of competition on price dispersion. Second, theories claim that the e ect of competition on the peak-load pricing based price dispersion should be similar between big city routes and tourist routes. But we find the results suggest otherwise. Our results in Table (16) reveal that overall competition would increase the price dispersion on routes with more heterogeneity in cross-price elasticity, like big city routes, while decrease the price dispersion on routes where travellers are more homogeneous in terms of cross-elasticity, like tourist routes. Even for the peak-load pricing based price dispersion, the effect of competition still differs in different types of routes and we find similar patterns between intertemporal price dispersion and across-date price dispersion, which suggests that there may be certain correlation between price discrimination and peak-loading pricing in reality although they are concepts from different theories. Last but not least, the cross-price elasticity between products is very sensitive to the definition of product. The cross-price elasticity can change greatly if we group a bunch of similar products slightly differently. For example, our results suggest that conditional on industry-elasticity not being dominant, heterogeneity in the cross-price elasticity among travellers is larger when we compare flights that only differ in their departure date and when we compare ticket purchases that only differ in the timing of buying the ticket. But cross-price elasticity is much more homogeneous if we look at flights that only differ in their take-off time. These conclusions clearly apply to other industries especially where consumers face big trade-off between inside and outside options and where there are a great deal of heterogeneity in consumers' preference over horizontal product differentiation.

Although we are cautious about interpreting our results as something causal, our finding does show that it is necessary to isolate price dispersion from different sources. We can see that increasing competition results in different trends in across-slot dispersion from across-date dispersion. The distinction attributes to the fact that different portions of price distribution respond to competition in different manners. Competition benefits consumers, who make early purchases or buy flights on non popular dates, to a greater degree than consumers who make late purchases or buy flights on popular dates.

But there is one caveat: Although the R-square of price level regression is high, the R-square is less than 0.2 in our baseline OLS[29] regressions of price dispersion[30], which implies we may miss some in uential factors, like tickets characteristics (Puller et al., 2008).

[29] Note that we don't report the R-square of the 2SLS regression which is also missing in Gerardi and Shapiro (2009) and Borenstein and Rose (1994), because the traditional R-square is not meaningful in the 2SLS case. Details can be seen in https://www.stata.com/support/faqs/statistics/two-stage-least-squares/.

[30] On average, our R-square in the regression of G^{IPD} is higher than that in Borenstein and Rose (1994). The R-square in the regression of G^{slot} is similar to that in Borenstein and Rose (1994).

References

Board, S. and Skrzypacz, A. (2016). Revenue management with forward-looking buyers. Journal of Political Economy 124(4):1046 1087.

Borenstein, S. (1985). Price discrimination in free-entry markets. The RAND Journal of Economics pages 380 397.

Borenstein, S. and Rose, N. L. (1994). Competition and price dispersion in the us airline industry. Journal of Political Economy 102(4):653 683.

Chandra, A. and Lederman, M. (2018). Revisiting the eelationship between competition and price discrimination. American Economic Journal: Microeconomics, 10(2):190 224.

Dai, M., Liu, Q., and Serfes, K. (2014). Is the e ect of competition on price dispersion nonmonotonic? evidence from the us airline industry. Review of Economics and Statistics 96(1):161 170.

Dana Jr, J. D. (1999). Equilibrium price dispersion under demand uncertainty: The roles of costly capacity and market structure. The RAND Journal of Economics, pages 632 660.

Dana Jr, J. D. and Williams, K. R. (2022). Intertemporal price discrimination in sequential quantity-price games. Marketing Science.

Eden, B. (1990). Marginal cost pricing when spot markets are complete. Journal of Political Economy 98(6):1293 1306.

Fang, H., Wang, L., and Yang, Y. (2020). Competition and quality: Evidence from high-speed railways and airlines. Technical report, National Bureau of Economic Research.

Gaggero, A. A. and Piga, C. A. (2011). Airline market power and intertemporal price dispersion. The Journal of Industrial Economics, 59(4):552 577.

Gale, I. L. and Holmes, T. J. (1993). Advance-purchase discounts and monopoly allocation of capacity. The American Economic Review, pages 135 146.

Gerardi, K. S. and Shapiro, A. H. (2009). Does competition reduce price dispersion? new evidence from the airline industry. Journal of Political Economy 117(1):1 37.

Goolsbee, A. and Syverson, C. (2008). How do incumbents respond to the threat of entry? evidence from the major airlines. The Quarterly journal of economics, 123(4):1611 1633.

Hernandez, M. A. and Wiggins, S. N. (2014). Nonlinear pricing strategies and competitive conditions in the airline industry. Economic Inquiry, 52(2):539 561.

Holmes, T. J. (1989). The e ects of third-degree price discrimination in oligopoly. The American Economic Review, 79(1):244 250.

Hortaçsu, A., Öry, A., and Williams, K. R. (2022). Dynamic price competition: Theory and evidence from airline markets.

Lazarev, J., Nevo, A., and Town, B. (2022). what can we learn from merger retrospectives? lessons for the u.s. airline industry.

Li, J., Granados, N., and Netessine, S. (2014). Are consumers strategic? structural estimation from the air-travel industry. *Management Science*, 60(9):2114 2137.

Prescott, E. C. (1975). E ciency of the natural rate. *Journal of Political Economy*, 83(6):1229 1236.

Puller, S. L., Sengupta, A., and Wiggins, S. N. (2008). Testing theories of scarcity pricing and price dispersion in the airline industry. *Worknig paper*.

Stavins, J. (2001). Price discrimination in the airline market: The e ect of market concentration. *Review of Economics and Statistics*, 83(1):200 202.

Stole, L. A. (2007). Price discrimination and competition. *Handbook of Industrial Organization*, 3:2221 2299.

Wang, J., Huang, J., and Jing, Y. (2020). Competition between high-speed trains and air travel in china: From a spatial to spatiotemporal perspective. *Transportation Research Part A: Policy and Practice*, 133:62 78.

Williams, K. R. (2022). The welfare e ects of dynamic pricing: Evidence from airline markets. *Econometrica*, 90(2):831 858.

Figure 1. Examples Of 3 Types Of Price Dispersion

(a) Over Booking Horizon

(b) Across Slot

(c) Across Departure Date

Figure 2. Load Factor By Departure Date, Departure Slot, and Booking Time

(a) By Departure Date

(b) By Slot

(c) By Booking Time

Table 1. Theory Prediction

Panel A: ($e^{industry}(p)$,$e^{cross\ price}(p)$)

	Tourist Routes	Big City Routes	Economic City Routes
Short-distance	(high, high)	(high,middle)	(high, low)
Medium-distance	(middle, high)	(middle,middle)	(middle, low)
Long-distance	(low, high)	(low,middle)	(low, low)

Panel B: Predicted e ect of competition on G^{IPD}

	Tourist Routes	Big City Routes	Economic City Routes
Short-distance	-	-	+
Medium-distance	-	?	+
Long-distance	-	?	+

Panel C: Predicted e ect of competition on G^{slot}

	Tourist Routes	Big City Routes	Economic City Routes
Short-distance	?	?	?
Medium-distance	?	?	?
Long-distance	?	?	?

Panel D: Predicted e ect of competition on G^{date}

	Tourist Routes	Big City Routes	Economic City Routes
Short-distance	?	?	?
Medium-distance	?	?	?
Long-distance	?	?	?

Table 2. Market Features By Market Structure

	Monopoly			Duopoly			Competitve		
	mean	sd	p50	mean	sd	p50	mean	sd	p50
Number Of Boarding Passengers	212.51	125.73	161.00	503.16	430.44	382.00	1180.67	827.36	848.00
Flight Distance	1370.18	709.85	1245.00	1378.89	628.65	1338.00	1403.55	417.08	1351.00
Number Of Big3 Carriers	1.00	0.00	1.00	1.46	0.50	1.00	2.01	0.70	2.00
Number Of Total Carriers							3.68	1.01	3.00
Observations	17168			29880			41334		

Table 3. Market Structure Summary

	mean	sd	p50
HHI	0.69	0.28	0.55
Number Of Carriers	1.91	1.03	2.00
Number Of Big3 Carriers	1.42	0.59	1.00
Weekly Flight Share Of Spring Airline	0.01	0.05	0.00
Observations	32406	32406	32406

Table 4. Gini Coe cient Summary

	G^{GS}	G^{IPD}	G^{slot}	G^{date}
min	0	0	0	0
max	0.384	0.394	0.402	0.376
mean	0.146	0.138	0.0949	0.183
sd	0.0642	0.0654	0.0782	0.0598
p50	0.143	0.134	0.0862	0.184
Observations	33352	56713	160466	12611

Table 5. Price Summary (Currency = RMB)

	Avg. Coach	Avg. Business	10th Percentile	90th Percentile
min	177.5	759	157.7	177.5
max	3177.8	6240	3069	4590
mean	820.3	2050.9	564.4	1324.3
sd	311.4	648.7	266.9	564.5
p50	774.0	1924	507.0	1260
Observations	33352	21456	33352	33352

Table 6. Median Gini-By Flight Distance By Market Structure

	Gerardi-Shapiro Gini			G^{IPD}			G^{slot}			G^{date}		
	Monopoly	Duopoly	Competitive	Monopoly	Duopoly	Competitive	Monopoly	Duopoly	Competitive	Monopoly	Duopoly	Competitive
less than 700km	0.124	0.144	0.149	0.130	0.132	0.138	0.0777	0.0806	0.0940	0.178	0.169	0.176
Observations	147	1319	1914	1783	2748	1602	683	6097	9099	250	633	525
700-1500km	0.133	0.151	0.140	0.133	0.135	0.129	0.0712	0.0850	0.0852	0.189	0.189	0.182
Observations	811	3198	12534	6228	9633	12532	3916	15486	59891	774	1945	3532
greater than 1500km	0.148	0.164	0.140	0.137	0.143	0.133	0.0782	0.102	0.0841	0.186	0.188	0.181
Observations	353	2977	10099	5174	7838	9175	1683	14618	48993	675	1620	2657

Table 7. Median Gini-By Route Type By Market Structure

	Gerardi-Shapiro Gini				G^{IPD}				G^{slot}				G^{date}			
	Big City	Tourist	Economic City	Whole Sample	Big City	Tourist	Economic City	Whole Sample	Big City	Tourist	Economic City	Whole Sample	Big City	Tourist	Economic City	Whole Sample
Monopoly	0.150	0.136	0.134	0.135	0.134	0.134	0.130	0.134	0.113	0.0701	0.0713	0.0738	0.187	0.186	0.182	0.186
Observations	277	809	654	1311	3971	8631	6450	13185	1348	3861	3135	6282	505	1085	795	1699
Duopoly	0.150	0.153	0.153	0.155	0.134	0.138	0.132	0.138	0.0947	0.0844	0.0903	0.0922	0.181	0.189	0.184	0.187
Observations	1924	5027	3127	7494	6786	11680	10226	20219	9245	24093	15060	36201	1256	2563	2119	4198
Competitive	0.143	0.139	0.139	0.140	0.132	0.132	0.126	0.131	0.0947	0.0743	0.0884	0.0852	0.181	0.184	0.176	0.181
Observations	10556	13867	14207	24547	12862	10806	15887	23309	50954	65941	68404	117983	3391	3479	4227	6714

Table 8. 2SLS Regression Of $\log(\text{Gini}^{IPD})_{ijst}$ On Competition: IV = log(scheduled capacity)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.201		0.005		0.114		0.518	
	(0.273)		(0.312)		(0.329)		(0.687)	
log(number of total carriers)		0.166		0.004		0.098		0.385
		(0.224)		(0.267)		(0.282)		(0.490)
Wald F Statistics	74.228	126.319	108.071	151.378	62.673	96.234	14.031	28.013
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.257		0.246		0.223		0.353	
	(0.093)		(0.098)		(0.133)		(0.142)	
log(number of total carriers)		0.215		0.210		0.176		0.283
		(0.077)		(0.081)		(0.101)		(0.111)
Wald F Statistics	245.124	407.611	122.997	187.699	72.347	144.687	115.469	217.523
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	-0.012		-0.018		-0.071		0.038	
	(0.068)		(0.094)		(0.151)		(0.080)	
log(number of total carriers)		-0.010		-0.016		-0.061		0.032
		(0.058)		(0.083)		(0.129)		(0.067)
Wald F Statistics	273.763	474.644	293.357	413.497	124.374	183.137	113.566	211.299
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.283		-0.148		-0.114		-0.444	
	(0.074)		(0.050)		(0.066)		(0.179)	
log(number of total carriers)		-0.234		-0.126		-0.098		-0.331
		(0.058)		(0.042)		(0.055)		(0.108)
Wald F Statistics	75.209	127.693	110.147	154.232	62.349	95.511	14.552	29.126
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.179		-0.179		-0.199		-0.218	
	(0.030)		(0.040)		(0.052)		(0.040)	
log(number of total carriers)		-0.150		-0.152		-0.157		-0.175
		(0.025)		(0.034)		(0.041)		(0.032)
Wald F Statistics	248.485	413.260	124.746	190.471	73.550	147.049	117.847	221.724
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.094		-0.067		-0.107		-0.121	
	(0.028)		(0.029)		(0.048)		(0.042)	
log(number of total carriers)		0.084		0.052		0.105		0.111
		(0.023)		(0.025)		(0.038)		(0.030)
Wald F Statistics	277.087	477.654	298.249	416.810	126.782	183.228	114.740	213.624
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.171		-0.161		-0.141		-0.318	
	(0.067)		(0.078)		(0.090)		(0.113)	
log(number of total carriers)		-0.141		-0.138		-0.121		-0.237
		(0.053)		(0.065)		(0.076)		(0.073)
Wald F Statistics	75.209	127.693	110.147	154.232	62.349	95.511	14.552	29.126
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.085		-0.113		-0.057		-0.101	
	(0.020)		(0.028)		(0.032)		(0.028)	
log(number of total carriers)		-0.071		-0.096		-0.045		-0.081
		(0.017)		(0.024)		(0.026)		(0.023)
Wald F Statistics	248.485	413.260	124.746	190.471	73.550	147.049	117.847	221.724
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	-0.075		-0.071		-0.106		-0.080	
	(0.020)		(0.025)		(0.033)		(0.026)	
log(number of total carriers)		-0.064		-0.063		-0.091		-0.067
		(0.017)		(0.022)		(0.028)		(0.021)
Wald F Statistics	277.087	477.654	298.249	416.810	126.782	183.228	114.740	213.624

Note: The details of xed e ect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

Table 9. 2SLS Regression Of $\log(\text{Gini}^{slot})_{ijbt}$ On Competition: IV = log(scheduled capacity)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	4.332		6.155		-0.262		11.701	
	(3.124)		(4.932)		(1.310)		(10.434)	
log(number of total carriers)		2.506		3.887		-0.124		8.109
		(1.601)		(2.593)		(0.611)		(5.863)
Wald F Statistics	7.197	19.727	4.345	10.970	1.060	5.119	2.759	6.561
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	-1.510		-2.078		-1.434		-1.278	
	(1.486)		(1.406)		(3.079)		(4.755)	
log(number of total carriers)		-1.201		-1.751		-0.927		8.109
		(1.175)		(1.168)		(1.970)		(5.863)
Wald F Statistics	17.484	27.357	25.032	35.345	4.406	11.548	3.910	6.561
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	-1.328		-1.252		-1.457		-3.222	
	(0.888)		(1.056)		(1.888)		(1.718)	
log(number of total carriers)		-0.903		-0.910		-0.835		-2.124
		(0.596)		(0.766)		(1.087)		(0.997)
Wald F Statistics	27.239	52.080	23.959	45.330	9.687	22.113	11.605	26.141
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.226		0.035		0.135		-0.106	
	(0.225)		(0.193)		(0.389)		(0.273)	
log(number of total carriers)		-0.134		0.023		0.070		-0.074
		(0.126)		(0.125)		(0.192)		(0.187)
Wald F Statistics	7.972	20.196	5.330	12.140	1.585	6.427	2.849	6.545
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	0.360		0.106		1.083		0.923	
	(0.218)		(0.113)		(0.769)		(0.779)	
log(number of total carriers)		0.284		0.089		0.698		0.552
		(0.164)		(0.094)		(0.394)		(0.395)
Wald F Statistics	17.275	27.550	23.997	34.427	4.392	11.561	3.804	14.221
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.339		-0.445		-0.196		-0.447	
	(0.128)		(0.168)		(0.239)		(0.239)	
log(number of total carriers)		-0.232		-0.324		-0.116		-0.296
		(0.082)		(0.112)		(0.139)		(0.144)
Wald F Statistics	28.729	53.956	24.827	45.938	10.718	23.743	12.040	26.404
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.180		-0.005		-0.142		0.069	
	(0.217)		(0.294)		(0.481)		(0.224)	
log(number of total carriers)		-0.107		-0.003		-0.073		0.048
		(0.133)		(0.190)		(0.270)		(0.156)
Wald F Statistics	7.972	20.196	5.330	12.140	1.585	6.427	2.849	6.545
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	0.384		0.138		1.150		1.232	
	(0.183)		(0.109)		(0.710)		(0.798)	
log(number of total carriers)		0.303		0.116		0.742		0.737
		(0.131)		(0.089)		(0.335)		(0.329)
Wald F Statistics	17.275	27.550	23.997	34.427	4.392	11.561	3.804	14.221
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	-0.197		-0.294		-0.267		-0.397	
	(0.137)		(0.183)		(0.285)		(0.261)	
log(number of total carriers)		-0.135		-0.214		-0.158		-0.263
		(0.093)		(0.131)		(0.166)		(0.161)
Wald F Statistics	28.729	53.956	24.827	45.938	10.718	23.743	12.040	26.404

Note: The details of xed e ect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

Table 10. 2SLS Regression Of $\log(Gini^{date})_{ijbs}$ On Competition: IV = log(scheduled capacity)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.780		1.001		-0.325		5.841	
	(0.782)		(1.342)		(0.442)		(5.468)	
log(number of total carriers)		0.712		1.104		-0.408		2.898
		(0.730)		(1.533)		(0.550)		(2.485)
Wald F Statistics	19.748	26.390	12.712	25.516	17.410	9.598	1.665	4.216
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.596		0.203		0.237		0.727	
	(0.329)		(0.203)		(0.102)		(0.513)	
log(number of total carriers)		0.608		0.193		0.221		2.898
		(0.334)		(0.194)		(0.098)		(2.485)
Wald F Statistics	104.394	101.787	61.985	62.819	48.281	58.088	63.436	4.216
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	0.877		1.287		2.139		0.247	
	(0.832)		(1.260)		(1.993)		(0.099)	
log(number of total carriers)		0.829		1.227		2.102		0.215
		(0.787)		(1.200)		(1.943)		(0.083)
Wald F Statistics	113.955	127.097	81.107	104.771	66.670	63.114	48.096	76.949
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.292		-0.145		-0.106		-1.149	
	(0.145)		(0.154)		(0.136)		(0.982)	
log(number of total carriers)		-0.269		-0.163		-0.134		-0.584
		(0.121)		(0.161)		(0.169)		(0.312)
Wald F Statistics	19.876	25.898	12.965	24.725	17.698	9.631	1.667	3.968
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.233		-0.221		-0.212		-0.246	
	(0.043)		(0.060)		(0.067)		(0.050)	
log(number of total carriers)		-0.237		-0.210		-0.197		-0.247
		(0.045)		(0.060)		(0.065)		(0.054)
Wald F Statistics	105.662	103.749	62.177	63.028	48.272	58.108	64.519	71.884
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.066		-0.048		-0.042		-0.079	
	(0.036)		(0.035)		(0.054)		(0.049)	
log(number of total carriers)		-0.062		-0.046		-0.042		-0.069
		(0.035)		(0.033)		(0.054)		(0.042)
Wald F Statistics	113.175	125.938	79.924	102.543	66.540	63.041	48.111	76.956
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.270		-0.440		-0.322		-0.456	
	(0.199)		(0.293)		(0.261)		(0.473)	
log(number of total carriers)		-0.248		-0.493		-0.405		-0.232
		(0.170)		(0.278)		(0.333)		(0.161)
Wald F Statistics	19.876	25.898	12.965	24.725	17.698	9.631	1.667	3.968
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.012		0.016		-0.041		-0.095	
	(0.031)		(0.044)		(0.048)		(0.035)	
log(number of total carriers)		-0.013		0.015		-0.038		-0.095
		(0.032)		(0.042)		(0.045)		(0.035)
Wald F Statistics	105.662	103.749	62.177	63.028	48.272	58.108	64.519	71.884
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	0.039		0.038		-0.028		0.089	
	(0.029)		(0.035)		(0.031)		(0.040)	
log(number of total carriers)		0.037		0.037		-0.028		0.077
		(0.027)		(0.033)		(0.031)		(0.033)
Wald F Statistics	113.175	125.938	79.924	102.543	66.540	63.041	48.111	76.956

Note: The details of fixed effect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Significant at the 10 percent significance level.
** Significant at the 5 percent significance level.
*** Significant at the 1 percent significance level.

Table 11. Regression Speci cation

Measure Of Competition	Variation	If Instrumented	Slot FE	Departure Date FE	Booking Time FE	Route-Carrier FE
		Panel A: G^{GS} and $P^{GS}(k)$				
-log(HHI)	cross-(route, departure date) variation	Y		Y		Y
log(number of total carriers)	cross-(route) variation	Y		Y		Y
		Panel B: G^{IPD} and $P^{IPD}(k)$				
-log(HHI)	cross-(route, slot, departure date) variation	Y	Y	Y		Y
log(number of total carriers)	cross-(route, slot) variation	Y	Y	Y		Y
		Panel C: G^{slot} and $P^{slot}(k)$				
-log(HHI)	cross-(route, departure date) variation	Y		Y	Y	Y
log(number of total carriers)	cross-(route) variation	Y		Y	Y	Y
		Panel D: G^{date} and $P^{date}(k)$				
-log(HHI)	cross-(route, slot) variation	Y	Y		Y	Y
log(number of total carriers)	cross-(route, slot) variation	Y	Y		Y	Y

Appendix

Appendix A

A.1 Replicating Gerardi and Shapiro (2009) Results

When replicating Gerardi and Shapiro (2009) with their instrument variable, Table (12) shows that more competition associates with higher price dispersion on big city routes. Low-end prices are more responsive to competition than high-end prices. Specifically, either a lower HHI or more carriers in the market would push up the 10th percentile price to a greater extent than the 90th percentile price. The sign of the competition's coefficients in the percentile price regression is the opposite to that in Gerardi and Shapiro (2009). This inconsistency may result from a bad application of their instrument in the context of China's airline industry, which we have discussed in previous sections. After using a novel instrument which we define in the main text, the effect of competition on the departure-date-route-level price dispersion, G^{GS}, is insignificant, we will argue that the insignificant results are the net outcome of different forces on different types of granular-level price dispersion. Another potential reason for the inconsistency with Gerardi and Shapiro (2009) is the difference in the aggregation level. Their data frequency is quarterly while our sample is daily, thus naturally price dispersion is larger as we define the market with a longer length of period.

Given the multi-dimension fixed effect terms in our setting, a consistent estimated coefficient can not be guaranteed when the sample size is small. In a small sample, the competition variable sometimes can be collinear with the fixed effects. Running our regression on subsamples within short routes is infeasible due to the small sample size. We are confined to run regression on the whole sample of short routes. For regressions with G^{date} being the dependent variable, getting the estimated coefficient of competition is also infeasible because of the collinearity issue. So we don't have regression results where G^{date} is the dependent variable.

For G^{IPD}, we find in Table (13) that competition has significant negative effect of price dispersion only on some types of routes, namely, short-distance tourist routes and all types of medium-distance routes. When the number of carriers increases, both the high-end 90th percentile price (i.e. price charged when booking is close to the departure) and the low-end 10th percentile price (i.e. price charged when booking is far from the departure) are lower, but to a different degree.

For G^{slot} in Table (14), competition has negative effect on price dispersion of long-distance routes while mixing effect on price dispersion of short- and medium-distance routes, but none of the coefficient of competition on price dispersion is significant. In terms of the effect on price distribution, more carriers will lower the high-end price (price of flights that have popular slot) on the medium-distance big city routes and long-distance tourist routes.

For G^{date} in Table (15), a less concentration market display higher price dispersion in general, but only the estimated efficient from short-distance economic city routes is significant. For long-distance routes, more competition correlates with higher price not only in the high-end segment (price of flights departing on popular date) but also in the low-end segment (price of flights departing on unpopular date).

Table 12. Gerardi and Shapiro (2009) 2SLS Regression, IV = log(enplanement)

	Whole Sample		Big City		Tourist		Economic City	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Y = Gini Coe cient Odd Ratio							
log(HHI)	0.122		0.753		-1.436		0.272	
	(0.339)		(0.378)		(1.142)		(0.621)	
log(number of total carriers)		0.101		0.646		-0.982		0.222
		(0.280)		(0.323)		(0.755)		(0.507)
	Panel B: Y = Price At 10th Percentile							
log(HHI)	0.319		0.110		1.254		0.500	
	(4.44)		(1.96)		(2.66)		(3.66)	
log(number of total carriers)		0.263		0.0941		0.855		0.407
		(4.70)		(2.00)		(3.58)		(3.99)
	Panel C: Y = Price At 90th Percentile							
log(HHI)	0.153		0.076		0.389		0.229	
	(0.057)		(0.057)		(0.200)		(0.089)	
log(number of total carriers)		0.126		0.065		0.265		0.187
		(0.047)		(0.049)		(0.121)		(0.070)

Note: hats indicate instrumental variable log(enplanement) is used. Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

Table 13. OLS Regression Of $\log(\text{Gini}^{IPD})_{ijst}$ On Competition

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	-0.061		-0.091		-0.002		-0.386	
	(0.113)		(0.143)		(0.156)		(0.149)	
log(number of total carriers)		-0.007		-0.044		0.041		-0.278
		(0.112)		(0.142)		(0.156)		(0.146)
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	-0.007		-0.032		-0.011		0.002	
	(0.050)		(0.046)		(0.060)		(0.073)	
log(number of total carriers)		0.025		-0.003		0.013		0.044
		(0.047)		(0.042)		(0.061)		(0.069)
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	0.034		0.052		0.103		0.024	
	(0.056)		(0.081)		(0.129)		(0.058)	
log(number of total carriers)		0.012		0.015		0.054		0.007
		(0.051)		(0.074)		(0.113)		(0.055)
Wald F Statistics	273.763	474.644	293.357	413.497	124.374	183.137	113.566	211.299
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.088		-0.058		-0.058		-0.031	
	(0.035)		(0.021)		(0.029)		(0.048)	
log(number of total carriers)		-0.106		-0.064		-0.067		-0.067
		(0.033)		(0.022)		(0.033)		(0.044)
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.065		-0.068		-0.073		-0.085	
	(0.017)		(0.020)		(0.026)		(0.022)	
log(number of total carriers)		-0.070		-0.075		-0.072		-0.087
		(0.016)		(0.018)		(0.024)		(0.020)
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.025		-0.029		-0.042		-0.025	
	(0.014)		(0.016)		(0.023)		(0.020)	
log(number of total carriers)		-0.025		-0.023		-0.034		-0.029
		(0.014)		(0.016)		(0.023)		(0.021)
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.067		-0.085		-0.076		-0.120	
	(0.029)		(0.039)		(0.046)		(0.038)	
log(number of total carriers)		-0.074		-0.087		-0.080		-0.133
		(0.028)		(0.038)		(0.045)		(0.032)
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.057		-0.086		-0.042		-0.069	
	(0.013)		(0.019)		(0.019)		(0.017)	
log(number of total carriers)		-0.055		-0.083		-0.039		-0.066
		(0.012)		(0.018)		(0.018)		(0.017)
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	-0.040		-0.052		-0.046		-0.028	
	(0.013)		(0.017)		(0.024)		(0.017)	
log(number of total carriers)		-0.042		-0.055		-0.046		-0.031
		(0.012)		(0.016)		(0.022)		(0.016)

Note: The details of xed e ect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

Table 14. OLS Regression Of $\log(\text{Gini}^{slot})_{ijbt}$ On Competition

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	-0.415		-0.690		0.307		-1.007	
	(0.530)		(0.721)		(0.296)		(1.037)	
log(number of total carriers)		0.218		0.292		0.314		-0.005
		(0.458)		(0.662)		(0.160)		(1.526)
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.037		0.177		-0.211		-0.082	
	(0.219)		(0.235)		(0.358)		(0.478)	
log(number of total carriers)		0.014		0.076		-0.076		-0.030
		(0.186)		(0.186)		(0.291)		(0.410)
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	-0.355		-0.329		-0.187		-0.287	
	(0.248)		(0.312)		(0.453)		(0.399)	
log(number of total carriers)		-0.212		-0.062		-0.004		-0.053
		(0.222)		(0.248)		(0.292)		(0.349)
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.045		-0.071		-0.129		-0.116	
	(0.050)		(0.065)		(0.130)		(0.103)	
log(number of total carriers)		-0.001		-0.006		-0.032		-0.009
		(0.037)		(0.033)		(0.056)		(0.051)
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.033		-0.033		-0.105		-0.038	
	(0.039)		(0.038)		(0.074)		(0.076)	
log(number of total carriers)		0.005		-0.005		-0.071		0.006
		(0.032)		(0.034)		(0.044)		(0.056)
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.030		-0.017		0.005		0.022	
	(0.035)		(0.049)		(0.056)		(0.059)	
log(number of total carriers)		-0.028		-0.019		0.014		-0.017
		(0.022)		(0.032)		(0.026)		(0.031)
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.068		-0.115		-0.209		-0.049	
	(0.079)		(0.105)		(0.170)		(0.062)	
log(number of total carriers)		0.005		-0.026		-0.082		0.032
		(0.101)		(0.131)		(0.201)		(0.074)
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.044		-0.042		-0.132		-0.063	
	(0.040)		(0.046)		(0.066)		(0.078)	
log(number of total carriers)		-0.009		-0.018		-0.075		-0.009
		(0.038)		(0.041)		(0.067)		(0.076)
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	-0.081		-0.099		-0.021		0.004	
	(0.034)		(0.046)		(0.046)		(0.046)	
log(number of total carriers)		-0.036		-0.044		0.029		0.034
		(0.028)		(0.037)		(0.026)		(0.034)

Note: The details of xed e ect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

Table 15. OLS Regression Of $\log(\text{Gini}^{date})_{ijbs}$ On Competition

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.190		0.280		-0.016		0.346	
	(0.147)		(0.242)		(0.162)		(0.389)	
log(number of total carriers)		-0.012		-0.086		-0.086		-0.161
		(0.108)		(0.163)		(0.133)		(0.206)
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.086		0.031		0.021		0.095	
	(0.067)		(0.083)		(0.044)		(0.103)	
log(number of total carriers)		0.108		0.005		0.019		0.123
		(0.086)		(0.090)		(0.042)		(0.134)
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	0.218		0.358		0.714		0.048	
	(0.250)		(0.403)		(0.706)		(0.032)	
log(number of total carriers)		0.138		0.221		0.616		0.042
		(0.225)		(0.339)		(0.622)		(0.034)
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.071		-0.041		-0.074		0.004	
	(0.042)		(0.045)		(0.053)		(0.078)	
log(number of total carriers)		-0.066		-0.018		-0.055		-0.059
		(0.043)		(0.064)		(0.053)		(0.077)
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.065		-0.059		-0.051		-0.086	
	(0.021)		(0.027)		(0.030)		(0.026)	
log(number of total carriers)		-0.062		-0.035		-0.038		-0.075
		(0.020)		(0.026)		(0.029)		(0.022)
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.007		-0.004		-0.021		-0.013	
	(0.017)		(0.020)		(0.031)		(0.022)	
log(number of total carriers)		0.008		0.017		-0.004		-0.007
		(0.017)		(0.018)		(0.035)		(0.024)
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.002		-0.032		-0.059		-0.033	
	(0.050)		(0.075)		(0.078)		(0.071)	
log(number of total carriers)		-0.057		-0.123		-0.092		-0.125
		(0.048)		(0.068)		(0.061)		(0.051)
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.026		-0.021		-0.041		-0.059	
	(0.016)		(0.026)		(0.025)		(0.019)	
log(number of total carriers)		-0.027		-0.021		-0.035		-0.061
		(0.015)		(0.023)		(0.025)		(0.019)
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	0.018		0.018		0.019		0.048	
	(0.017)		(0.021)		(0.024)		(0.024)	
log(number of total carriers)		0.012		0.010		0.021		0.045
		(0.017)		(0.022)		(0.022)		(0.023)

Note: The details of xed e ect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Signi cant at the 10 percent signi cance level.
** Signi cant at the 5 percent signi cance level.
*** Signi cant at the 1 percent signi cance level.

A.2 Using Pre-Scheduled Capacity as IV

We estimate the regressions in Gerardi and Shapiro (2009) but with the logarithm of pre-scheduled capacity as instrument this time. As we can see in (16), here we can see that reduction in market concentration will lead to drop in price, which is more intuitive than the results in (12). And we don't see a significant impact of competition on the price dispersion at a more aggregated level.

Table 16. Gerardi and Shapiro (2009) 2SLS Regression, IV = log(scheduled capacity)

	Whole Sample		Big City		Tourist		Economic City	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Y = Gini Coefficient Odd Ratio							
$-log(HHI)$	0.049		0.279		-0.343		0.330	
	(0.252)		(0.284)		(0.577)		(0.459)	
log(number of total carriers)		0.040		0.234		-0.236		0.271
		(0.204)		(0.239)		(0.397)		(0.379)
	Panel B: Y = Price At 10th Percentile							
$-log(HHI)$	-0.0694		-0.0997*		0.0710		-0.0643	
	(-1.75)		(-2.58)		(0.51)		(-1.10)	
log(number of total carriers)		-0.0561		-0.0837**		0.0489		-0.0527
		(-1.76)		(-2.60)		(0.52)		(-1.10)
	Panel C: Y = Price At 90th Percentile							
$-log(HHI)$	-0.175***		-0.202***		-0.136		-0.115**	
	(0.043)		(0.044)		(0.133)		(0.055)	
log(number of total carriers)		-0.142***		-0.170***		-0.093		-0.094**
		(0.034)		(0.036)		(0.090)		(0.045)

Note: hats indicate instrumental variable log(enplanement) is used. Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Significant at the 10 percent significance level.
** Significant at the 5 percent significance level.
*** Significant at the 1 percent significance level.

Appendix B

Using the logarithm of enplanement as instrument for competition, we estimate the equations from (1) to (12). In Table (18), the instrument fails to pass the weak instrument test and the results that more competition would push up prices seems counter-intuitive.

Table 17. 2SLS Regression Of $log(Gini^{IPD})_{ijst}$ On Competition: IV = log(enplanement)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.018		-0.044		0.082		0.289	
	(0.230)		(0.264)		(0.263)		(0.532)	
log(number of total carriers)		0.015		-0.039		0.074		0.217
		(0.194)		(0.232)		(0.237)		(0.390)
Wald F Statistics	83.739	133.272	115.507	140.413	89.589	105.334	16.401	32.421
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.116		0.203**		-0.063		0.150	
	(0.095)		(0.097)		(0.133)		(0.143)	
log(number of total carriers)		0.098		0.175**		-0.051		0.122
		(0.080)		(0.084)		(0.107)		(0.115)
Wald F Statistics	282.945	430.229	156.672	205.946	87.482	157.184	133.824	228.266
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	-0.015		0.029		-0.065		-0.036	
	(0.070)		(0.095)		(0.154)		(0.090)	
log(number of total carriers)		-0.013		0.026		-0.056		-0.030
		(0.060)		(0.085)		(0.133)		(0.076)
Wald F Statistics	276.066	433.908	302.535	409.813	117.421	169.675	115.438	206.732
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.165***		-0.057		-0.022		-0.300**	
	(0.061)		(0.043)		(0.054)		(0.133)	
log(number of total carriers)		-0.140***		-0.051		-0.020		-0.226**
		(0.051)		(0.037)		(0.048)		(0.087)
Wald F Statistics	85.504	135.637	118.602	145.147	89.058	104.816	16.984	33.517
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.114***		-0.134***		-0.108**		-0.136***	
	(0.030)		(0.040)		(0.054)		(0.041)	
log(number of total carriers)		-0.096***		-0.116***		-0.087**		-0.110***
		(0.026)		(0.034)		(0.043)		(0.033)
Wald F Statistics	286.542	435.886	158.750	209.219	88.671	159.463	136.368	232.580
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.054**		-0.036		-0.057		-0.063	
	(0.027)		(0.029)		(0.049)		(0.040)	
log(number of total carriers)		-0.046**		-0.032		-0.050		-0.053
		(0.023)		(0.026)		(0.043)		(0.034)
Wald F Statistics	279.100	434.433	308.243	409.253	119.491	168.145	116.622	208.916
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.083		-0.052		-0.058		-0.226**	
	(0.068)		(0.081)		(0.090)		(0.098)	
log(number of total carriers)		-0.070		-0.046		-0.052		-0.170**
		(0.056)		(0.071)		(0.081)		(0.067)
Wald F Statistics	85.504	135.637	118.602	145.147	89.058	104.816	16.984	33.517
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.043**		-0.066**		-0.025		-0.059**	
	(0.021)		(0.029)		(0.031)		(0.028)	
log(number of total carriers)		-0.036**		-0.057**		-0.020		-0.048**
		(0.017)		(0.025)		(0.025)		(0.023)
Wald F Statistics	286.542	435.886	158.750	209.219	88.671	159.463	136.368	232.580
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	-0.032		-0.026		-0.072**		-0.038	
	(0.020)		(0.025)		(0.033)		(0.027)	
log(number of total carriers)		-0.028		-0.023		-0.063**		-0.032
		(0.017)		(0.022)		(0.028)		(0.023)
Wald F Statistics	279.100	434.433	308.243	409.253	119.491	168.145	116.622	208.916

Note: The details of fixed effect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Significant at the 10 percent significance level.
** Significant at the 5 percent significance level.
*** Significant at the 1 percent significance level.

Table 18. 2SLS Regression Of $log(Gini^{slot})_{ijbt}$ On Competition: IV = log(enplanement)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.699		5.654		-2.954		8.890	
	(2.609)		(5.180)		(4.142)		(11.754)	
log(number of total carriers)		0.435		3.285		-1.957		5.155
		(1.631)		(2.596)		(2.321)		(5.373)
Wald F Statistics	5.991	12.503	3.124	9.635	2.011	3.621	1.240	4.846
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	-5.412*		-3.214		-19.097		-24.999	
	(2.858)		(2.148)		(26.033)		(35.436)	
log(number of total carriers)		-3.885**		-2.524		-6.125		5.155
		(1.887)		(1.649)		(4.426)		(5.373)
Wald F Statistics	9.572	19.907	11.272	19.103	0.615	7.352	0.538	4.846
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	-1.812		-1.073		-3.667		-3.452	
	(1.475)		(1.622)		(3.261)		(2.558)	
log(number of total carriers)		-1.242		-0.813		-2.251		-2.571
		(0.999)		(1.227)		(1.929)		(1.805)
Wald F Statistics	21.226	36.351	21.105	31.084	8.483	15.325	9.650	14.133
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	1.369*		1.150		1.334		1.968	
	(0.715)		(0.823)		(1.393)		(1.836)	
log(number of total carriers)		0.863**		0.697*		0.861		1.198
		(0.346)		(0.391)		(0.707)		(0.705)
Wald F Statistics	6.451	12.984	3.932	10.463	2.299	4.293	1.482	5.014
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	2.241**		0.673**		10.046		16.267	
	(0.894)		(0.315)		(14.723)		(25.513)	
log(number of total carriers)		1.586***		0.522**		2.992**		4.049**
		(0.478)		(0.216)		(1.361)		(1.550)
Wald F Statistics	9.094	19.619	10.566	18.501	0.494	6.889	0.420	9.017
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	0.410		0.065		0.886		0.894*	
	(0.253)		(0.225)		(0.618)		(0.534)	
log(number of total carriers)		0.283*		0.049		0.563		0.666*
		(0.169)		(0.170)		(0.369)		(0.379)
Wald F Statistics	22.477	37.879	21.974	32.153	9.453	16.242	9.925	14.368
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	1.089*		1.195		1.131		1.614	
	(0.599)		(0.861)		(1.207)		(1.297)	
log(number of total carriers)		0.687**		0.724*		0.730		0.982**
		(0.310)		(0.400)		(0.633)		(0.445)
Wald F Statistics	6.451	12.984	3.932	10.463	2.299	4.293	1.482	5.014
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	1.513**		0.598**		6.405		10.924	
	(0.602)		(0.287)		(9.439)		(17.040)	
log(number of total carriers)		1.071***		0.463**		1.907**		2.719***
		(0.320)		(0.195)		(0.879)		(1.007)
Wald F Statistics	9.094	19.619	10.566	18.501	0.494	6.889	0.420	9.017
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	0.463**		0.243		0.297		0.459	
	(0.223)		(0.233)		(0.319)		(0.353)	
log(number of total carriers)		0.319**		0.184		0.189		0.342
		(0.147)		(0.176)		(0.198)		(0.259)
Wald F Statistics	22.477	37.879	21.974	32.153	9.453	16.242	9.925	14.368

Note: The details of fixed effect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Significant at the 10 percent significance level.
** Significant at the 5 percent significance level.
*** Significant at the 1 percent significance level.

Table 19. 2SLS Regression Of $log(Gini^{date})_{ijbs}$ On Competition: IV = log(enplanement)

	All Routes		Tourist Routes		Big City Routes		Economic City Routes	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A-1: Short Routes, Y = log(Gini)								
-log(HHI)	0.648		0.783		-0.218		4.528	
	(0.688)		(1.095)		(0.389)		(3.944)	
log(number of total carriers)		0.601		0.852		-0.274		2.538
		(0.654)		(1.232)		(0.480)		(2.232)
Wald F Statistics	25.891	33.069	17.464	33.562	32.720	16.259	2.439	4.694
Panel A-2: Medium Routes, Y = log(Gini)								
-log(HHI)	0.586*		0.190		0.227**		0.724	
	(0.331)		(0.194)		(0.100)		(0.520)	
log(number of total carriers)		0.608*		0.184		0.215**		2.538
		(0.342)		(0.189)		(0.098)		(2.232)
Wald F Statistics	110.551	101.148	72.566	66.214	56.534	62.047	67.759	4.694
Panel A-3: Long Routes, Y = log(Gini)								
-log(HHI)	0.893		1.313		2.213		0.249**	
	(0.846)		(1.280)		(2.046)		(0.100)	
log(number of total carriers)		0.853		1.262		2.215		0.220**
		(0.807)		(1.228)		(2.024)		(0.085)
Wald F Statistics	115.446	123.663	81.010	101.812	62.825	60.093	47.897	74.343
Panel B-1: Short Routes, Y = log(10th Price)								
-log(HHI)	-0.234*		-0.098		-0.086		-0.913	
	(0.120)		(0.116)		(0.106)		(0.681)	
log(number of total carriers)		-0.218**		-0.107		-0.109		-0.522*
		(0.102)		(0.121)		(0.130)		(0.264)
Wald F Statistics	26.092	32.559	17.775	32.832	33.256	16.304	2.436	4.458
Panel B-2: Medium Routes, Y = log(10th Price)								
-log(HHI)	-0.230***		-0.212***		-0.209***		-0.248***	
	(0.042)		(0.059)		(0.066)		(0.049)	
log(number of total carriers)		-0.238***		-0.206***		-0.198***		-0.253***
		(0.045)		(0.060)		(0.065)		(0.053)
Wald F Statistics	111.650	103.163	72.774	66.446	56.516	62.072	68.666	71.974
Panel B-3: Long Routes, Y = log(10th Price)								
-log(HHI)	-0.072**		-0.055		-0.050		-0.082	
	(0.037)		(0.036)		(0.056)		(0.050)	
log(number of total carriers)		-0.069*		-0.053		-0.050		-0.073*
		(0.035)		(0.034)		(0.057)		(0.043)
Wald F Statistics	114.478	122.314	79.554	99.172	62.647	59.958	47.910	74.351
Panel C-1: Short Routes, Y = log(90th Price)								
-log(HHI)	-0.241		-0.364		-0.269		-0.427	
	(0.178)		(0.241)		(0.214)		(0.396)	
log(number of total carriers)		-0.225		-0.400*		-0.339		-0.244
		(0.155)		(0.232)		(0.265)		(0.156)
Wald F Statistics	26.092	32.559	17.775	32.832	33.256	16.304	2.436	4.458
Panel C-2: Medium Routes, Y = log(90th Price)								
-log(HHI)	-0.012		0.016		-0.042		-0.097***	
	(0.031)		(0.043)		(0.048)		(0.034)	
log(number of total carriers)		-0.012		0.016		-0.040		-0.099***
		(0.032)		(0.042)		(0.045)		(0.035)
Wald F Statistics	111.650	103.163	72.774	66.446	56.516	62.072	68.666	71.974
Panel C-3: Long Routes, Y = log(90th Price)								
-log(HHI)	0.032		0.033		-0.035		0.084**	
	(0.029)		(0.036)		(0.032)		(0.040)	
log(number of total carriers)		0.031		0.031		-0.035		0.074**
		(0.028)		(0.034)		(0.033)		(0.034)
Wald F Statistics	114.478	122.314	79.554	99.172	62.647	59.958	47.910	74.351

Note: The details of fixed effect please refer to Table (11). Standard error is clustered at the route level. N in the table above stands for the number of observations.
* Significant at the 10 percent significance level.
** Significant at the 5 percent significance level.
*** Significant at the 1 percent significance level.