informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 10, October 2023, pp. 6076–6108
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Managing Airfares Under Competition: Insights from a Field Experiment

Maxime C. Cohen,[a,*] Alexandre Jacquillat,[b] Juan Camilo Serpa,[a] Michael Benborhoum[c]

[a] Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada; [b] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [c] Twilio, New York, New York

*Corresponding author

Contact: maxime.cohen@mcgill.ca, https://orcid.org/0000-0002-2474-3875 (MCC); alexjacq@mit.edu (AJ); juan.serpa@mcgill.ca (JCS); michael.benborhoum@gmail.com (MB)

Received: July 19, 2020
Revised: June 29, 2021; May 6, 2022; November 18, 2022
Accepted: November 28, 2022
Published Online in Articles in Advance: January 31, 2023

https://doi.org/10.1287/mnsc.2022.4656

Abstract. Airfares evolve dynamically, giving rise to a so-called *price path*. This price path is controlled via two levers: (i) a fare ladder, which defines a set of airfares before the selling season, and (ii) revenue management algorithms, which control how fares evolve along the ladder during the season. We hypothesize that the current policies to control both levers—which do not account for quality differences between competing airlines—give rise to an inefficient price path and, accordingly, a loss of potential revenue. We substantiate this hypothesis via a field experiment. By partnering with an airline, we introduced quality considerations in the design of fare ladders, across 5,000 itineraries, to show that current ladder-design policies indeed lead to a suboptimal price path. We also show that this inefficiency can be mitigated by incorporating quality differences between competing airlines. This creates a smoother (and more profitable) price path.

History: Accepted by Vishal Gaur, operations management.

Keywords: airline pricing • field experiment • lead-in fare

1. Introduction

Airlines change their fares dynamically. To determine these fares, they first create a fare ladder for each itinerary. This ladder includes N fare classes $\{C_1, C_2, \ldots, C_N\}$, which are known as "inventory classes," with corresponding fares $\{p_1, \ldots, p_N\}$, where $p_i \le p_j$ for $i \le j$. The fare of the lowest inventory class, p_1, is called the *lead-in fare*, whereas $\{p_2, \ldots, p_N\}$ are called *sell-up* fares.

Once this ladder is designed, the selling season begins. Throughout the season, the displayed fare evolves gradually as a function of the number of bookings, the time remaining to departure, and other relevant factors. Let $F_t = p_i^*(t)$ be the fare displayed at time t that naturally corresponds to the fare of inventory class C_i^*. To set F_t, airlines rely on revenue management algorithms, which operate in one of two ways: (i) via quantity controls (i.e., by letting F_t be the fare of the lowest inventory class whose sales are below some predetermined target) or alternatively, (ii) via bid price controls (i.e., by letting F_t be the smallest fare p_i such that $p_i \ge b_t$, for some pre-established b_t).

The evolution of F_t throughout the selling season is called a *price path*. In Figure 1(a), we illustrate a hypothetical ladder with nine inventory classes, and in Figure 1(b), we depict a possible price path for this ladder. In this exhibit, the fare changed on five occasions (at t_1, t_2, t_3, t_4, and t_5), jumping from p_1 to p_2, and eventually rising to p_7.

Passengers are not aware of the underlying fare ladder or the revenue management algorithms; they only see the displayed fare, F_t, at any given time. However, because F_t ultimately determines the passengers' booking decisions, it is essential to control the price path to maximize the airline's revenue.

Airlines can alter the price paths in two ways: (i) by modifying the design of the fare ladder, ahead of the selling season; or (ii) by modifying revenue management algorithms that govern the evolution of airfares during the selling season. Our goal in this paper is to identify (and alleviate) inefficiencies in the airfare management systems by focusing on the first approach, namely, by studying the design of the fare ladder.

1.1. How Do Airlines Design the Fare Ladder?

To design the fare ladder, airlines do not manually set every fare. Instead, they anchor the sell-up fares to the lead-in fare. In Figure 1, for example, the airline would only choose a value for p_1, and the sell-ups would then be determined via prespecified functions (such that $p_2 = f_2(p_1)$, $p_3 = f_3(p_1)$, etc.). It is precisely for this reason that the lead-in fare receives its name: because it shapes the rest of the ladder.

Currently, competitors match each other's lead-in fares. And because airlines use almost-identical (if not the same) functions to price their sell-ups, they end up competing with virtually the same ladder. For example,

Figure 1. (Color online) Airfare Ladder with Nine Inventory Classes and a Hypothetical Price Path

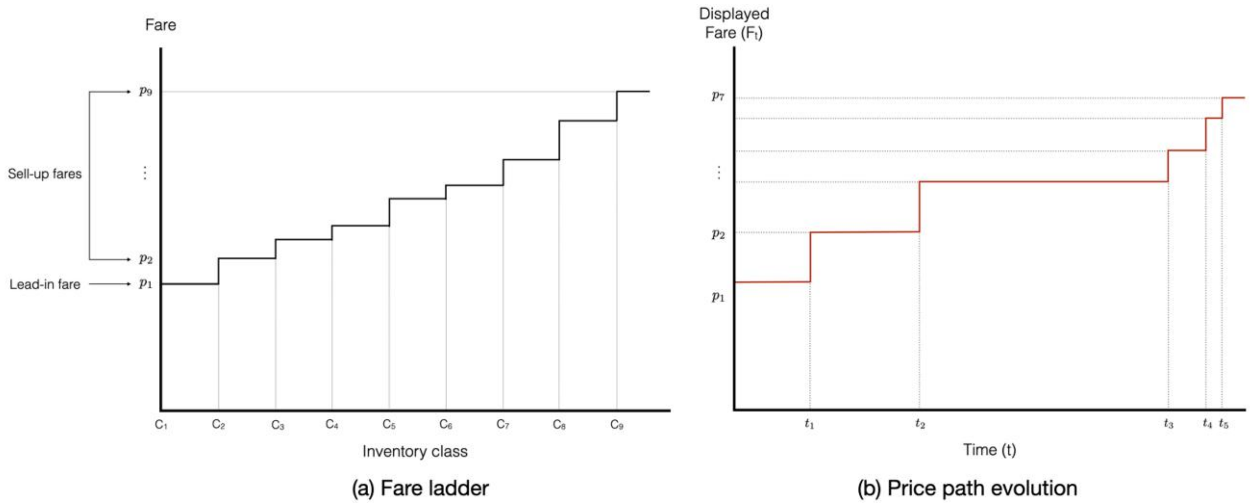


Figure 2 shows the ladders of three airlines (American Airlines, Delta Airlines, and United Airlines) for the {Austin–Seattle} itinerary departing June 22, 2020, and returning June 25, 2020. The three airlines have the same lead-in fare ($168) and a virtually identical ladder with 25 inventory classes for the economy cabin. This practice, known as *lead-in fare matching*, is the norm and not the exception. After surveying the ladders of every airline across 9,000 markets worldwide, we found that competitors match their ladders across 96.5% of the itineraries (see details in Section 3.2).

Airlines knowingly decide to compete with the same fare ladders even when there are significant differences in service quality—for instance, a direct versus a connecting itinerary, a short versus a long layover, or a convenient versus an inopportune departure time. Figure 3 shows two examples from itineraries that had just been released (and that, at the time, were both displaying the lead-in fare). In the first example, Air France and Royal Air Maroc offered the same fare for the {Montreal–Toulouse} itinerary, even though Royal Air Maroc's itinerary is 27 hours longer. In the second example, Air Canada and its three U.S. competitors were offering the same fare for the {Toronto–Orlando} itinerary, although the Canadian company offers the only direct flight.

Why would airlines match their ladders when there are patent differences in itinerary quality? Rather than using the fare ladder to proactively differentiate fares, competitors rely on the revenue management algorithms to do so. If an airline offers a better itinerary, consumers will book it more aggressively, the displayed fares will rise faster and, accordingly, the price path of a high-quality itinerary will be steeper than that of a low-quality itinerary. In this way, the revenue management algorithms ensure that better itineraries will display a higher fare, even if they have the same underlying ladder. We corroborated this fact by looking at Air France's fare for the two previous examples: the {Montreal–Toulouse} itinerary had climbed to US$1,194 within a few days, whereas Royal Air Maroc's fare only increased to US$831. And Air Canada's fare ended up outpacing that of its competitors.

1.2. Research Questions

Although airlines have two levers to control the price path—the fare ladder and the revenue management algorithms—these levers are interdependent. Changing the design of the fare ladder may affect the efficiency of the revenue management algorithms, and vice versa. Airline pricing research, however, has overwhelmingly attempted to optimize the revenue management algorithms by taking the fare ladder as a given input—effectively overlooking the interaction between the two levers. By examining two questions, we aim to bridge this gap:

1. Under the current ladder-design policy—which is based on lead-in fare matching—are the existing revenue management algorithms giving rise to an efficient price path?

2. If not, what mechanisms can alter the evolution of the price path to maximize revenue?

1.3. Hypotheses

When it comes to the first research question, we hypothesize that the answer is *no*, especially for high-quality itineraries. We conjecture that under the current policy, in which airlines match their ladders irrespective of quality differences, the revenue management dynamics will be inefficient and, accordingly, produce suboptimal price paths. This hypothesis is illustrated in the appendix with a stylized model that emulates the

Figure 2. (Color online) Ladders of Three Airlines for {Austin–Seattle} Itineraries Departing June 22, 2020, and Returning June 25, 2020

Search: Departing AUS on 06/22/20 for SEA
Returning on 06/25/20
Flying AA UA DL
Viewing in currency USD
Searching in Economy cabin
Validated Results

Show fares for cabins:

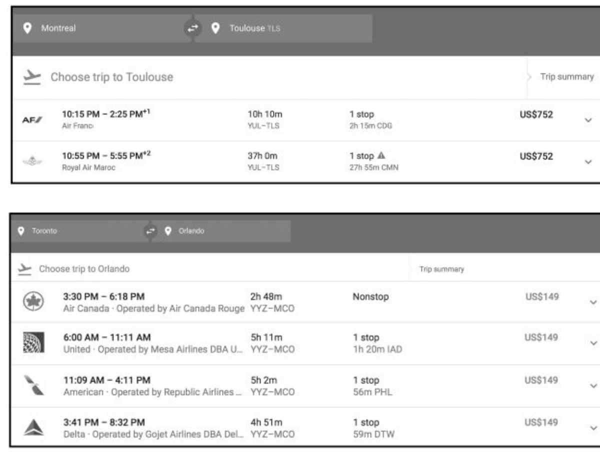
Fare Basis	Airline	Booking Class	Trip Type	Fare	Cabin
GAA2AQBN	UA	G	Round-Trip	168.00(USD)	E
GRA2AQER	UA	G	Round-Trip	168.00(USD)	E
OVAKZNB5	AA	O	Round-Trip	168.00(USD)	E
OVAKZNO5	AA	O	Round-Trip	168.00(USD)	E
XAVQA0BC	DL	X	Round-Trip	168.00(USD)	E
XAVQA0EE	DL	X	Round-Trip	168.00(USD)	E
TAVNA0BC	DL	T	Round-Trip	208.00(USD)	E
TAVNA0EE	DL	T	Round-Trip	208.00(USD)	E
TAVOA0BC	DL	T	Round-Trip	208.00(USD)	E
TAVOA0EE	DL	T	Round-Trip	208.00(USD)	E
GAA4AQBN	UA	G	Round-Trip	228.00(USD)	E
GRA4AQER	UA	G	Round-Trip	228.00(USD)	E
OVAKZNM5	AA	O	Round-Trip	228.00(USD)	E
QUAKZNB5	AA	Q	Round-Trip	228.00(USD)	E
QUAKZNO5	AA	Q	Round-Trip	228.00(USD)	E
TAUQA0BC	DL	T	Round-Trip	228.00(USD)	E
TAUQA0EE	DL	T	Round-Trip	228.00(USD)	E
XAVQA0ME	DL	X	Round-Trip	228.00(USD)	E
GAA2AQDN	UA	G	Round-Trip	238.00(USD)	E
KAA2AQBN	UA	K	Round-Trip	268.00(USD)	E
KRA2AQER	UA	K	Round-Trip	268.00(USD)	E
TAVNA0ME	DL	T	Round-Trip	268.00(USD)	E
TAVOA0ME	DL	T	Round-Trip	268.00(USD)	E
TAVUA0BC	DL	T	Round-Trip	268.00(USD)	E
TAVUA0EE	DL	T	Round-Trip	268.00(USD)	E
KAA4AQBN	UA	K	Round-Trip	288.00(USD)	E
KRA4AQER	UA	K	Round-Trip	288.00(USD)	E
NUAKZNB5	AA	N	Round-Trip	288.00(USD)	E
NUAKZNO5	AA	N	Round-Trip	288.00(USD)	E
QUAKZNM5	AA	Q	Round-Trip	288.00(USD)	E
TAUQA0ME	DL	T	Round-Trip	288.00(USD)	E

Source. Airline Tariff Publishing Company (ATPCO).
Notes. AA, American Airlines; DL, Delta Airlines; UA, United Airlines. In this exhibit, each airline offers two itineraries on this departure–return schedule; we only report a subset of the ladders due to space constraints (in the entire ladder, the airlines matched the fares in 23 of 26 inventory classes). The query date for both ladders is February 18, 2020.

following dynamics. Suppose we have two airlines competing over a given trip: one airline has a desirable itinerary (e.g., a direct flight), whereas the other has an undesirable one (e.g., a trip with multiple layovers). In this scenario, our model shows that when the two airlines match their ladders, the high-quality itinerary will end up with suboptimal fares during the early part of the selling season and, eventually, yield suboptimal revenue.

When it comes to the second research question, we argue that airlines can mitigate such inefficiencies by adopting an alternative policy, which we call *quality-based fare management*. Specifically, we suggest that airlines should mark up the lead-in fare of high-quality itineraries, that is, that they should design a marked-up ladder. This strategy will smooth out the price path

Figure 3. Displayed Fares for Two Itineraries



Source. Google Flights.
Note. Query date: April 2018.

throughout the selling season, meaning that the airline will sell fewer tickets early on but more toward the end of the season when prices are higher. The result highlights a dynamic substitution in sales: even though the airline loses revenue early on by marking up the ladder, in the end its revenue increases more than it would otherwise.[1]

1.4. Methodology and Results

To substantiate our hypotheses, we partnered with one of the world's leading airlines and implemented a large-scale field experiment. In this experiment, we marked-up the lead-in fares—and, hence, the ladders—of 4,913 "high-quality" itineraries across 22 origin-destination markets, randomly selected from a large pool of candidates (where quality is defined via the Quality Service Index, an industry metric that ranks itineraries based on nonpricing attributes). In other words, we replaced the existing lead-in matching policy for a quality-based fare management approach—namely, by implementing a markup on the lead-in fare.

In total, our experiment affected 86,246 bookings. We also randomly selected 13,842 itineraries, belonging to 121 markets, to serve as our control group, i.e., which continued to operate under lead-in matching. To deploy this experiment, we collaborated with a revenue management team, which monitored all 18,755 itineraries (across 143 markets) during the experiment. The airline avoided using special promotions during the experiment, which could have interfered with our treatment. By running this experiment, we show evidence for our hypotheses by providing three insights:

R1 (Baseline result). In Section 7.1, we show that under the current lead-in fare matching policy, the revenue management algorithms give rise to an inefficient price path for high-quality itineraries. Such inefficiencies can be mitigated by marking up the ladder, that is, by implementing quality-based fare management. Specifically, we establish that marking up the ladder of a high-quality itinerary—as opposed to matching it—increases revenue and yield (by 0.26 and 0.23 standard deviations, respectively) without sacrificing sales and market share.

R2 (Mechanism). In Section 7.2, we demonstrate that the baseline result is driven by a smoother price path. To this end, we dissect our results across the experiment's timeline to study how the price path evolves under both pricing strategies. Particularly, this analysis shows that, by implementing up the ladder, the airline replaces early (low-valued) sales with late (high-valued) sales. This dynamic substitution in sales causes a smoother price path through the selling season and, thereby, increases revenues.

R3 (Boundaries). In Section 7.3, we show that the performance edge of the quality-based fare management strategy is bounded by four key market conditions. Specifically, this section identifies factors that make ladder markups most (or least) beneficial. We find that our proposed strategy performs better with high-traffic destinations, popular travel hours, long travel times, and competitive markets. Elsewhere, the profitability edge disappears. We also explain the economic intuition behind these results and prescribe how airlines could leverage the two pricing strategies to optimize revenue.

1.5. Contribution

Today, airfare management systems are implemented via a two-step procedure that involves (i) designing a fare ladder and (ii) developing revenue management algorithms that control the fare evolution through capacity controls or bid price controls. Together, these two steps determine the price path throughout the selling season.

Currently, managers operationalize this two-step procedure via two joint practices: ladder matching (i.e., when designing their price path) and ex post differentiation (i.e., when designing the revenue management algorithms). As we show in our experiment, this status quo leaves unrealized revenue on the table. As such, our paper exposes inefficiencies in the current interaction of airline pricing and revenue management algorithms.

Our paper identifies revenue-enhancing opportunities for airlines, namely, through quality-based fare management. On high-quality itineraries, airlines can induce a smoother price path that shifts sales toward the later part of the selling season. In the current distribution environment, this strategy can be implemented through two distinct avenues. The first suggestion is that airlines can improve the architecture of the fare ladder (ex ante) without impacting revenue management dynamics (ex post). Simply put, rather than systematically matching their competitors' lead-in fares, airlines could integrate quality differences into the design of the fare ladder itself. The second suggestion is that managers could revise their revenue management algorithms—even under a ladder-matching strategy—to strengthen ex post pricing differentiation under competition, based on itinerary quality.

Moving forward, the interaction between these two practices will lie at the core of airline pricing as the industry transitions from traditional class-based revenue management toward more granular fare classes and continuous pricing. With this transition come new complexities in the design of pricing and revenue management algorithms. Our results provide interpretable guidelines to improve the design of next-generation pricing and revenue management algorithms. Our experiment can also serve to identify markets and itineraries where ongoing practices are most inefficient, and thus where new algorithms are most warranted.

2. Literature Review

Belobaba (1989) proposes an approach to dynamically allocate the number of seats in each inventory class for a given fare ladder. This approach became a standard in airline revenue management, inspiring many follow-up studies, such as those surveyed by McGill and Van Ryzin (1999), Bitran and Caldentey (2003), and Talluri and Van Ryzin (2006). Most closely related to our paper are Feng and Gallego (1995) and Bitran and Caldentey (2003), who use dynamic programming to optimize revenue management as a function of demand and available inventory.

Revenue management researchers have since developed more complex models and algorithms that capture network effects and customer behavior. For instance, Gallego and Van Ryzin (1994) integrate dynamic pricing and inventory management across networks. Talluri and Van Ryzin (1998) consider bundled products (e.g., multileg itineraries, multinight hotel stays). Talluri and Van Ryzin (2004) incorporate discrete-choice models that account for customers' buy-ups and buy-downs (i.e., when customers substitute between low and high fares). These complexities make dynamic revenue management computationally challenging. To tackle this problem, Zhang and Adelman (2009) and Zhang and Weatherford (2017) develop efficient algorithms for large data sets.

Airline revenue management becomes more intricate in multifirm environments. Using simulations, Belobaba and Wilson (1997) find that dynamic revenue management is still beneficial in a competitive setting—especially for the airline that adopts a revenue management system first. Netessine and Shumsky (2005) model an airline inventory management game with two players, finding that competition may induce higher prices. Perakis and Sood (2006) and Adida and Perakis (2010) solve dynamic pricing models under competition by using robust optimization with variational inequalities and fluid models.

In a competitive environment, airline pricing ultimately hinges on how customers make booking decisions. Empirically, Koppelman et al. (2008) and Lurkin et al. (2017) fit discrete-choice models—using historical booking data—to estimate the sensitivity of passengers to airfares and itinerary characteristics (e.g., trip duration, departure time). Martínez-de Albéniz and Talluri (2011) and Gallego and Hu (2014) integrate discrete-choice models into a game-theoretical framework for revenue management. Martínez-de-Albéniz and Talluri (2011) use Bertrand competition to show that the firm with the highest reservation price will set a fare equal to that of its competitors. This result suggests that lead-in fare matching can sometimes be optimal. Gallego and Hu (2014), however, find that price matching can be undesirable when firms offer differentiated products.

Methodologically, our paper contributes to the empirical literature on retail pricing. Within this literature, Olivares and Cachon (2009) use cross-sectional variation in sales and inventory to identify the impact of the entry and exit decisions by competitors in the automotive industry. Also using data from the automotive industry, Moreno and Terwiesch (2015) characterize the relationship between product mix flexibility and pricing decisions.

Our paper is also related to the empirical literature on airline management. Ramdas et al. (2013) investigate the effect of operational performance (e.g., on-time performance, lost bags, and denied boardings) on stock returns. Li et al. (2014) use a structural model that incorporates air travelers' behavior and find that 5.2%–19.2% of the customers are strategic when purchasing tickets. Nicolae et al. (2016) quantify the impact of free-checked-bag policies on operational performance.

Unlike these studies, we run a controlled field experiment. Gaur and Fisher (2005) design a field experiment by clustering "similar" treatment-control stores to estimate the relationship between prices and sales. Caro and Gallien (2012) design a pricing model for fast-fashion retailers and test it on Zara's stores, showing a potential $200–300 million revenue increase. Fisher et al. (2018) design a dynamic pricing model to determine how retailers should respond to competitors' price changes and partner with an online retailer to validate their model with a field experiment. Our paper also tests pricing under competition but focuses on whether an airline should *proactively* deviate from lead-in matching rather than *reactively* respond to competitors' price changes.

3. Background

3.1. Competition and Quality in the Airline Industry

Retailers compete across *markets* by offering *products*; in contrast, airlines compete across origin-destinations (*ODs*) by offering *itineraries*.

An OD is defined by a pair of cities, for example, the {Boston–Seattle} OD. Each OD represents a market, and if an airline decides to compete in a given OD, it does so by offering itineraries. An itinerary is defined as a trip, with a given schedule, from the origin to the destination airport. An itinerary could serve the OD via a direct flight or a set of connecting flights. For example, a {Paris–Istanbul} trip on September 14 and a {Paris–Istanbul} trip on September 15 are two different itineraries within the same OD.

Some airlines offer a single weekly itinerary in a given OD, whereas others offer upward of 10 daily itineraries. The quality of these itineraries depends on the number of connections, the layover time, and the departure time (among other factors). To measure these traits, airlines use the quality service index (QSI), a well-established industry metric that ranks itineraries, within an OD, based on objective and measurable nonpricing attributes like the ones mentioned previously (Welch 2012, Belobaba et al. 2015).

3.2. Building the Fare Ladder

Airlines create a different fare ladder for each itinerary by choosing the number of inventory classes and then assigning a fare to each. To this end, they use a *lead-in pricing* strategy, which anchors the upper fares to the bottom-most fare, namely, the "lead-in fare."

For instance, in Figure 1, the airline assigned nine inventory classes, with corresponding fares $(p_1, p_2, \ldots, p_9)$. In this ladder, the airline would only choose p_1, whereas the sell-ups would be determined via prespecified pricing rules specific to each itinerary, such that $p_2 = f_2(p_1)$, $p_3 = f_3(p_1)$, $p_4 = f_4(p_1)$, etc. Once the ladder is in place, the selling season begins and, as seats get booked (or as time elapses), the revenue management systems move the displayed fare up or down the ladder, that is, to control the price paths.[2]

The sell-up functions are tailored to each OD and even to each itinerary. Although airlines keep their sell-up functions secret, the data show that competitors use almost-identical functions. This happens because they share the same pricing solutions providers or because, after decades of practice, they have converged to an industry equilibrium.

3.2.1. How Often Do Airlines Match Their Lead-in Fares?

During site visits to our airline partner, we witnessed, firsthand, that most markets were embroiled in this practice. However, although lead-in fare matching is ubiquitous in the industry, there are unfortunately no public statistics about this practice. To fill this void, we gathered data on fare ladders. All airlines disclose their entire fare ladders to the Airline Tariff Publishing Company (ATPCO). By purchasing a subscription, we were able to query the ladders of all airlines across all itineraries worldwide that were scheduled for the upcoming year (the screenshots from Figure 2 show results from two queries).

With our subscription, we randomly sampled 9,000 itineraries across 30 days. To this end, we considered the world's 100 busiest airports, based on passenger traffic, and queried itineraries from this list by randomly selecting (i) origin and destination airports and (ii) departure and return dates. We then collected the fare ladders of all airlines catering each of these itineraries. Collectively, we gathered 3,085,703 fares, spread across 176,901 airfare ladders, 8,831 itineraries, and 204 airlines (169 queries returned corrupt or unreadable data, so we discarded them from the sample).

We found that lead-in matching happens in 8,518 of the 8,831 sampled itineraries—that is, in 96.46% of the cases. In these 8,518 markets, the median market included nine airlines, and eight of these airlines practice lead-in fare matching. When such matching occurs, 91.5% of the inventory classes in the economy cabin are matched. In other words, airlines that match their lead-in fares end up having identical ladders. The reason for having a 91.5% overlap—as opposed to a 100% overlap—is that some airlines create additional inventory classes to offer discounts to special segments (e.g., for veterans, the military, or corporate partners). Barring these differences, it is safe to say that airlines price the sell-ups using the same algorithms. In our random sample, 195 (of 204) airlines match their lead-in fares in at least one itinerary and, on average, 78% of the competitors match their lead-in fares in every market. The size of the "matching coalition" increases with the number of competitors, almost linearly.

We highlight that there is no particular subgroup of airlines that match their lead-in fares, or that airlines strategically decide when to match and when not to match. Instead, lead-in matching is an almost-universal practice, with virtually every airline matching their ladders. Most often, the airlines that shy away from matching are *very* small players (e.g., local airlines specializing in an OD) or specialty carriers that compete in restricted markets (e.g., private jet carriers).

3.3. Quality Differentiation: The Role of the Revenue Management Algorithms

Intuitively, the lead-in fare should pivot as a function of itinerary quality (e.g., better departure times, more legroom space, fewer number of connections, better in-flight entertainment). Accordingly, higher-quality itineraries should have a higher ladder. In reality, however, airlines disregard quality when choosing the lead-in fare and instead choose to match their competitors' lead-in fares. This means that most airlines end up with the same lead-in fare and, by extension, with the same ladder, although there may be large differences in itinerary quality.

To understand why lead-in fare matching has been ubiquitously practiced for decades, we interviewed pricing managers in the industry, who impressed on us their notion that in the airline industry, consumer demand—not airlines—should drive fare differentiation. If an itinerary is truly better (due, for instance, to a more attractive scheduled time, fewer connections, or bundled ancillary services), customers will book it more often, and accordingly, the revenue management algorithms will move up the ladder more quickly, leading to a steeper price path. Thus, quality is not integrated into the ladder's architecture but left to the revenue management algorithms, which adjust the price path based on demand until the market reaches a natural "fare differential."

To illustrate, let $F_{H,t}$ and $F_{L,t}$ be the displayed fares at time t for a high-quality itinerary and its low-quality counterpart. Because $F_{H,t}$ and $F_{L,t}$ are determined via quantity-based or bid-price controls—both of which are tied to demand—then we would expect the fares of the high-quality itineraries to rise faster at the beginning of the season, that is, that $\Delta_t F_{H,t} > \Delta_t F_{L,t}$, when both itineraries have been launched and are still at the lead-in price. As time progresses, the displayed fares $F_{H,t}$ and $F_{L,t}$ will begin to diverge until they reach an equilibrium price differential $(F_{H,t} - F_{L,t})$ that reflects their true quality differential (at least according to consumer demand).

3.4. Research Questions

Despite the wealth of research in revenue management, this research has taken the design of the ladder as a given input. There is, however, scant research about how the design of the fare ladder affects the optimality of the revenue management systems.

In response, this paper addresses the following two research questions. Under the current ladder-design policy—which is based on lead-in fare matching—are the existing revenue management algorithms giving rise to an efficient price path? If not, which mechanisms can be used to alter the evolution of observed fares and increase revenues?

Answering these questions will allow us to highlight inefficiencies in the existing fare management systems and propose mechanisms to alleviate such inefficiencies.

3.5. Hypothesis

We hypothesize that the current policies for designing ladders—which are based on lead-in fare matching—create inefficiencies in the revenue management dynamics, particularly because the current fare management systems do not account for quality differences in the itineraries of competitors. As a result, we argue that high-quality itineraries will display suboptimally low fares early in the selling season, resulting in a loss in potential revenue for the airline. We also hypothesize that the airline will benefit by altering the price path through quality-based fare management, either by adjusting fare ladders, based on quality, or by modifying revenue management methods.

The intuition is that by matching the ladder of a high-quality itinerary, the airline provides an early "bargain" for consumers. Thus, as consumers rush to buy the high-quality itinerary at the outset, this creates a larger share of bookings early on. This initial burst of sales will make the price path steeper in the early part of the selling season but flatter in the late part, meaning that the airline will sell fewer tickets later in the season (precisely because customers book most seats at the outset). In contrast, when the airline marks up the ladder of a high-quality itinerary, the revenue management algorithms will produce a smoother price path. The airline will thus sell fewer tickets early on, relative to the price path of a matched ladder, causing the fares to rise at a steadier pace but more toward the end of the season, when ticket prices are higher. We root this hypothesis on a stylized model, which we develop in the appendix.

4. Field Experiment

To test our hypothesis, we deployed a large-scale field experiment by partnering with an airline. Our partner is a top 10 legacy airline that competes across thousands of ODs by offering millions of itineraries yearly.

We needed to clear some obstacles before running our field experiment. In an ideal world, we would run an A/B test by randomly displaying different fares to different customers. A booker would see a fare that stems either out of a matching strategy or a markup strategy. We would then compare the booker's behavior under both strategies. Unfortunately, this design is infeasible because airlines cannot display different fares to different customers for the same itinerary at the same time. However, we can design an experiment that is close to ideal by taking advantage of the large set of markets served by the airline. To this end, we engineered a new kind of experiment for airline research, which uses a multidimensional control group and a treatment-switch approach. This experimental design controls for both temporal and cross-sectional variation across observations. Specifically, we follow seven steps.

4.1. Step 1. Designing the Treatment: Quality-Based Fare Management

We will modify the ladder design, moving away from lead-in fare matching. Instead, we propose a new approach, which we call quality-based fare management, which marks up the ladders of high-quality itineraries. To test this alternative strategy, we randomly choose a set of ODs (from a large pool) and then "treat" itineraries within these ODs. By "treating" an itinerary, we mean that we mark up its ladder by proactively raising the lead-in fare above the matching equilibrium. We treat all itineraries within each OD, including round-trip and one-way itineraries. To illustrate our treatment, consider Figure 4. Suppose that inventory classes $C_1, \ldots, C_7$ belong to the economy cabin, whereas C_8 and C_9 are in the business cabin. In the status quo scenario, our airline partner would have a ladder with fares $\{p_1, \ldots, p_9\}$, where p_1 is matched with the competition (and, by extension, so are the sell-up fares). Our treatment consists of an upward shock to the lead-in, α_1, and a bundle of shocks, $\{\alpha_2, \alpha_3, \ldots, \alpha_7\}$, to the nonpremium sell-ups. The values of $\alpha_2, \ldots, \alpha_7$ are determined by the sell-up functions, such that $p_2 + \alpha_2 = f_2(p_1 + \alpha_1), \ldots, p_7 + \alpha_7 = f_7(p_1 + \alpha_1)$. Accordingly, "treated" itineraries display fares from a ladder designed via quality-based fare management, whereas "controlled" itineraries display fares from a matched ladder. We then compare how sales perform across these itineraries. When the lead-in fare (p_1) is no longer the displayed fare, the treatment still impacts the remaining sell-up fares in the ladder. Suppose, for instance, that the airline has closed up Classes C_1 and C_2. Under a matching policy, the airline sells the remaining seats at prices p_3, p_4, and so on. Under a markup policy, the airline sells the remaining seats at prices $p_3 + \alpha_3$, $p_4 + \alpha_4$, and so on.

To understand why we chose this specific treatment, recall that airlines have three pricing levers: (i) the lead-in fare, (ii) the sell-up functions, and (iii) the revenue management dynamics. Our treatment reflects only the impact of marking up the lead-in fare, while leaving the other two levers intact.

4.2. Step 2. Choosing the Markup Level

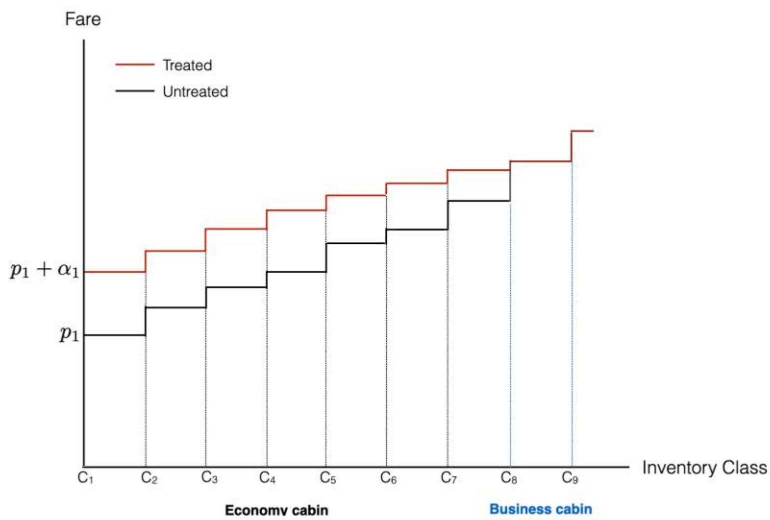
To choose the markup level, we first decided to apply the same percentage markup across all treated itineraries. Having a uniform markup allows us to estimate treatment effects consistently. Also, our goal is to determine how a reasonable differential in the ladder affects sales performance (determining the optimal markup is beyond the paper's scope).

In terms of the actual markup, Proposition C.2 (in our stylized model) warns us against using a markup level that is too high or too low. And this makes sense intuitively: A 50% markup will drive all customers away, whereas a 0.5% markup level will drive away bargain hunters (without being sufficiently high to make up for the loss of customers). Thus, we decided to apply a midpoint markup, that is, a markup that is not too low (thereby ensuring that consumers notice a nontrivial change) but also not too high (thereby minimizing an overblown markup level). We discussed options that ranged from 0.5% to 8%. After receiving advice from the airline pricing managers, we agreed that a 5% markup was an appropriate midpoint markup.

From the ATPCO fare data, a 5% increase in the lead-in fare represents a \$112.20 increase for the average fare of an international round-trip itinerary. The change, of course, varies along the ladder: at the bottom, a 5% markup is close to a \$60 increase; in the upper portion, the increase is closer to \$195. Furthermore, these differences

Figure 4. (Color online) Treatment Effect

Inventory Class	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Untreated ladders	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9
Treated ladders	$p_1+\alpha_1$	$p_2+\alpha_2$	$p_3+\alpha_3$	$p_4+\alpha_4$	$p_5+\alpha_5$	$p_6+\alpha_6$	$p_7+\alpha_7$	p_8	p_9



often stack up because 41% of the bookings include multiple tickets—Thus, for someone booking a family trip, a 5% markup in the lead-in fare could easily represent a difference of US$250–$400.

To be clear, our treatment does not necessarily imply that the displayed fare will always be 5% higher than that of the competitors. It just means that the ladder itself will be raised (relative to the competitors' ladders). In other words, we are exclusively modifying the structure of the ladder; the fares displayed depend on how the revenue management dynamics play out.

Another clarification is that a 5% markup on the lead-in fare does not mean that each step of the ladder will be raised *exactly* by 5% because, recall, the shock $\alpha = 5\%$ is infused inside the sell-up functions (see Step 1). In the aggregate, however, the resulting markup oscillates between 5% and 5.5% across all inventory classes in the ladder—that is, consistently close to the 5% benchmark.

4.3. Step 3. Selecting the Experimental Population

Before randomly selecting a treatment group and a control group, we need to define the experimental population. Given the nature of the experiment, we are not interested in studying all itineraries or all ODs. Instead, we restrict our attention to an experimental pool that satisfies five conditions:

1. *High-quality itineraries.* We focus on itineraries that outpace competitors in terms of quality—it would not make sense to mark up the ladder of a low-quality itinerary. To this end, we use a normalized measure of the QSI that ranges between zero and one, such that the sum of all airlines' QSIs equals one. Our experimental pool—both in the treatment and control groups—will only contain itineraries for which the partner airline's QSI exceeds 0.5. This ensures that these itineraries are of higher quality than the competition. In a given OD market, there is little within-airline variation when it comes to itinerary quality. This is because an airline often uses the same aircraft, legroom space, departing time, and in-flight service on a given OD. Moreover, our experimental population is restricted to ODs where all itineraries offered by the airline are of higher quality than those offered by its competitors. All itineraries of the tested ODs are characterized by a high normalized QSI.

2. *Itineraries that are still at the bottom of the ladder.* It would not be sound to study the ladder of an itinerary that departs in two days, or one that is entirely booked. For this reason, we focus on itineraries that are still at the bottom of the ladder—and thus with most seats still available—at the beginning of the trial.

3. *Itineraries with enough potential demand.* We do not want to experiment on itineraries that, say, depart in 15 months; these itineraries would receive one or two bookings, irrespective of prices. In fact, our data show that 89% of the bookings happen within four months of departure. For this reason, we focus on itineraries that are within 120 days of the departure date, a condition that allows us to capture itineraries during the "hot part" of the selling season. We also eliminate ODs for

which the airline only offers one monthly itinerary, or with remote destinations that only include a dozen weekly bookings. We decided to treat itineraries with different departure dates to avoid the influence of time-variant idiosyncrasies. If all itineraries had been departing, say, on August 1, we could have confounded the effect of our experiment with effects that are idiosyncratic to this date. For this reason, we chose a staggered approach, which includes a continuum of itineraries with different departure dates.

4. *ODs managed by our local revenue management team.* Airlines divide ODs across revenue management offices, which are scattered across the world. It would be prohibitively expensive and inefficient to run the experiment across multiple locations, since we would have to be physically present in different cities. Thus, we focus on ODs managed by the U.S. offices (note this team manages tens of thousands of ODs, meaning that this is not a stringent constraint).

5. *ODs catering international, long-haul flights with at least one connection.* This restriction allows us to enhance the experiment, given that nonstop itineraries are idiosyncratic and, accordingly, difficult to benchmark against a control group. In contrast, multistop itineraries often have good control itineraries: {Houston–Paris–Marseille}, for example, can be reliably paired with {Houston–Paris–Nice} and {Dallas–Paris–Marseille}. This is not a strong restriction either because our airline is a major international player that offers hundreds of thousands of itineraries with these traits.

We also exclude from the experimental pool a few ODs that could not be treated due to external constraints—for example, because of commercial agreements or legal restrictions in the industry.

4.4. Step 4. Selecting the Treated ODs

The five restrictions allow us to gather a clean and reliable experimental population. This pool includes tens of thousands of ODs and hundreds of thousands of itineraries. Using this pool, we randomly select the treated ODs.

In particular, the airline agreed to treat 22 ODs for eight weeks. We selected these 22 ODs via a random number generator. This list ended up yielding 4,913 itineraries, which were varied in terms of geography, size, and traffic.

4.5. Step 5. Selecting the Control ODs

We created a multidimensional control group with a treatment-switch approach, to absorb several streams of unobserved heterogeneity and, thus, obtain reliable results. In particular, we included three types of control groups: a counterpart control group, a twin control group, and a reverse control group (see Table 1 for an illustration):

1. *Counterpart control group.* We partition our 22 treated ODs into two subsets of 11 ODs: test set 1, denoted by $TS_1 = \{OD_1^1, \ldots, OD_{11}^1\}$; and test set 2, denoted by $TS_2 = \{OD_1^2, \ldots, OD_{11}^2\}$. From Week 1 to Week 4, we apply the treatment to TS_1 while controlling TS_2. Then, from Week 5 to Week 8, we apply the treatment to TS_2 while controlling TS_1.

2. *Twin control group.* We create a twin control group, $Twin = \{OD_1^T, \ldots, OD_{88}^T\}$. Each treated OD has a set of four "twin" ODs. Itineraries in the twin ODs are almost identical to itineraries in the treated ODs (in terms of geography, trip time, departure time, layover time, demand, and QSI). We define $Twin_1$ as the set of twin ODs for TS_1 and $Twin_2$ as the set of twin ODs for TS_2. None of the itineraries from these ODs were treated. The twin control group was carefully curated in collaboration with airline managers who confirmed that this is an appropriate design for an experiment at the OD level in the airline industry.

3. *Reverse control group.* We create the set $Reverse = \{OD_1^R, \ldots, OD_{22}^R\}$ of the 22 ODs with reverse origins and destinations. For instance, if the {Vancouver– Bogota} OD is treated, the {Bogota–Vancouver} OD is in the reverse

Table 1. Treatment and Control Groups

	Twin	Reverse
Test set 1		
{Austin–Los Angeles—Stuttgart}	{Dallas–Los Angeles—Stuttgart}	{Stuttgart–Los Angeles—Austin}
	{Houston–Los Angeles—Stuttgart}	
	{Austin–Los Angeles—Munich}	
{Edmonton–Amsterdam—Liverpool}	{Edmonton–Amsterdam—Manchester}	{Liverpool–Amsterdam—Edmonton}
	{Edmonton–Amsterdam—Liverpool}	
	{Edmonton–Amsterdam—Leeds}	
Test set 2	Twin	Reverse
{Montreal–Barcelona—Malaga}	{Montreal–Barcelona—Seville}	{Malaga–Barcelona—Montreal}
	{Montreal–Barcelona—Murcia}	
	{Montreal–Barcelona—Granada}	
{Denver–Paris—Cologne}	{Denver–Paris—Dortmund}	{Cologne–Paris—Denver}
	{Denver–Paris—Essen}	
	{Denver–Paris—Muenster}	

Notes. This sample is for illustrative purposes only but resembles our actual sample. Our sample comprises 11 ODs in each test group.

set. Each itinerary has one reverse itinerary. We define $Reverse_1$ as the set of reverse ODs for TS_1 and $Reverse_2$ as the set of reverse ODs for TS_2, respectively. None of the itineraries of the reverse ODs was treated. This group is a good benchmark for three reasons. First, airlines provide the same service level on an itinerary and its reverse itinerary. Thus, the QSI values are almost identical (and, for the same reason, so are the competitors' QSIs). Second, an itinerary and its reverse have nearly identical competitive environments because the OD and its reverse are served by the same airlines. Third, price fluctuations are similar between an OD and its reverse. Although price tags might not be the same, at any given point, price trends are highly correlated in an itinerary and its reverse. Accordingly, any deviation in the trends would be a strong indicator of a significant treatment effect.

Although any control group above remains imperfect on its own, these three different control groups offer different dimensions of similarity and allow us to provide a wholesome counterfactual group to enhance the validity of our experimental design. As Table 2 shows, ODs in TS_1 were treated and ODs in $\{TS_2, Twin, Reverse\}$ were not treated during Weeks 1–4; in turn, ODs in TS_2 were treated and ODs in $\{TS_1, Twin, Reverse\}$ were not treated during Weeks 5–8.

4.6. Step 6. Defining the Outcome Variables

Our goal is to measure sales performance. To this end, we selected the following four metrics:

- *Revenue*: The income from all bookings, measured from the airline's data.
- *Yield*: The ratio of revenue to the number of tickets sold, measured from the airline's data.
- *Tickets sold*: The number of seats booked, measured from the airline's data.
- *Market share*: The fraction of tickets sold within the OD, measured from the Computer Reservation System.

We aggregate these metrics at the OD-week level and normalize by expressing them in standard deviations relative to the deseasonalized historical average. We also add dummy variables to account for trend effects and to control for weekend and holiday effects. Ultimately, we measure the airline's expected bookings and gauge the performance against the demand expectation for the week. This normalization allows us to circumvent issues related to seasonality, capacity, or demand. For instance, some itineraries have larger sales on weekends or during specific weeks—or may simply have a larger demand inflow. By measuring abnormal demand, relative to average historical data, we eliminate these idiosyncratic differences.

Finally, when computing these four metrics, we do not include sales from premium classes (e.g., in Figure 4 we would only use Classes C_1 to C_7 to compute the outcome variables). This is because the treatment does not directly affect the premium classes, and the demand for economy and premium classes are largely disjoint. For robustness, however, we re-estimated our results by including premium classes (and our results were unaffected).

Table 2. Experimental Design

Group	Weeks 1–4	Weeks 5–8
Test set 1	Treatment	Control
Test set 2	Control	Treatment
Twin	Control	Control
Reverse	Control	Control

4.7. Step 7. Deploying and Monitoring the Experiment

To deploy our experiment, we collaborated with a revenue management team that helped us monitor all daily treated and control itineraries. The team also monitored the competitors' fares. Specifically, we discussed the possibility that competitors would rematch the ladders. If this were to happen, we agreed to raise the fares again to always maintain a 5% lead-in markup. However, this situation never happened: the competitors did not deviate from the matching equilibrium on any itinerary.

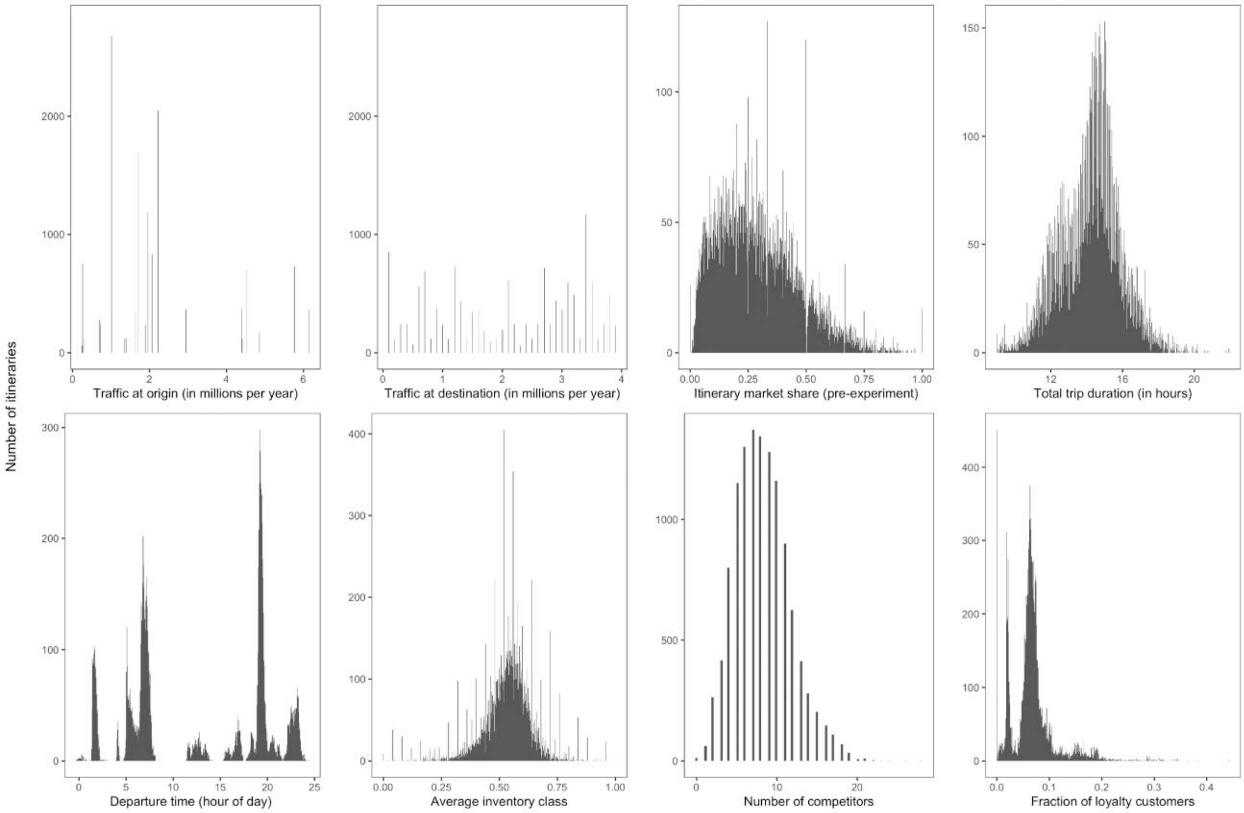
During the trial, the revenue management team also agreed to keep the revenue management algorithms unaltered, that is, by opening and closing inventory classes just as they had been doing before the experiment. The airline also avoided special promotional events both across the treatment and control groups. By doing so, we isolated the experiment from potential confounding factors.

5. Raw Statistics

To determine how similar the treatment and control groups are, we gather eight key characteristics about the sampled itineraries:

1. *Traffic at the origin.* Annual number of departures, in millions, at the origin airport (measured at the OD level).
2. *Traffic at the destination.* Annual number of departures, in millions, at the destination airport (measured at the OD level).
3. *Pre-experiment market share.* The airline's market share in the year before our experiment (measured at the OD-week level).
4. *Total trip duration.* The itinerary's duration from origin to destination, in hours, including all flights and layovers (measured at the itinerary level).
5. *Departure time.* Local departure time of the itinerary's first flight (measured at the itinerary level).
6. *Average inventory class.* This continuous variable—which spans from zero to one—captures how the ladder advanced across itineraries in the previous year (measured at the OD-week level). When the metric is close to zero, most sales stemmed from low inventory

Figure 5. Distribution of Control Variables Across Itineraries



classes; and when it is close to one, most sales came from high inventory classes.

7. *Number of competitors.* Number of competitors, including our partner, that also offered an itinerary on the same OD and on the same day (measured at the OD-day level).

Figure 5 plots the variables' distribution for the aggregate sample, which reveals a high degree of variability for the eight variables across the sampled itineraries. Table 3 reports the variables' mean and standard deviation, disaggregated by group. From the table, we can see that the four groups are highly balanced. In fact, a test of population differences, in the distributions' first and second moments, fails to reject the null hypothesis across all these variables. In other words, the four groups are statistically indistinguishable from each other, in mean and variance, across all control variables.

In terms of the outcome variables, Table 4 shows that, before the experiment, all groups performed close to their historical average. However, these figures change

Table 3. Summary Statistics: Control Variables

	TS_1		TS_2		Twin		Reverse	
Variable	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Traffic at the origin	1.81	1.38	2.08	1.50	1.89	1.45	1.95	1.70
Traffic at the destination	1.76	1.48	2.27	1.89	1.92	1.35	1.98	1.10
Pre-experiment market share	0.32	0.16	0.32	0.16	0.28	0.14	0.28	0.17
Total trip duration	14.81	4.05	14.71	5.79	14.76	5.80	14.32	5.04
Departure time (hour of day)	12.77	6.35	12.12	5.45	12.24	4.29	12.85	5.98
Average inventory class	0.51	0.21	0.48	0.25	0.53	0.19	0.49	0.20
Number of competitors	8.36	2.77	8.86	2.41	8.26	3.01	8.96	2.63
Itineraries	2,422		2,491		8,807		5,035	

Table 4. Summary Statistics: Outcome Variables

		TS_1		TS_2		Twin		Reverse	
	Variable	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Pre-experiment	*Revenue*	−0.02	0.97	−0.02	0.97	−0.03	1.00	0.00	1.00
	Yield	0.00	0.99	0.00	1.00	−0.01	1.03	0.02	1.01
	Tickets sold	0.09	0.80	−0.10	0.84	−0.08	0.91	0.11	0.80
	Market share	0.01	1.01	0.01	1.006	0.00	1.00	0.02	1.01
Weeks 1–4	*Revenue*	0.44	1.42	−0.04	0.73	0.17	1.11	0.11	0.97
	Yield	0.41	1.16	−0.33	0.81	−0.08	0.92	−0.18	0.78
	Tickets sold	−0.02	0.84	0.35	1.47	0.14	1.17	0.12	1.28
	Market share	−0.19	0.75	−0.25	0.88	−0.04	0.93	−0.30	0.76
Weeks 5–8	*Revenue*	0.10	1.10	0.90	1.74	0.49	1.45	0.06	1.05
	Yield	−0.26	0.75	0.38	1.29	0.03	1.04	-0.29	0.73
	Tickets sold	0.31	1.35	−0.06	0.89	0.25	1.05	0.33	1.23
	Market share	−0.16	0.74	−0.13	0.95	−0.05	1.03	-0.19	0.86

Notes. The variables are measured in standard deviations relative to the historical deseasonalized average. The variables *Revenue, Yield, Tickets Sold,* and *Market Share* have been standardized to have a standard deviation of one.

during the experimental trial: Table 4 shows that, during Weeks 1–4, the revenue and yield of TS_1's itineraries jumped by 0.44 and 0.41 standard deviation. Similarly, during Weeks 5–8, the revenue and yield of TS_2's itineraries increased by 0.9 and 0.38 standard deviation. Moreover, the effect on market share and sales was weaker. Figure 6, which plots the probability distribution functions of the outcome variables, confirms the impact of our experiment on revenue and yield: the distribution of these two metrics spreads out toward the right. In contrast, the treatment did not meaningfully affect the distribution of sales and market share.

6. Identification

We identify the treatment effect by running two econometric models: (i) cross-group comparisons and (ii) difference-in-differences. These two tests benchmark each other. The first one identifies the treatment effect by relying on cross-sectional variation, whereas the second one relies on variation in the time series, that is, by analyzing trend changes. Here, we explain these two identification approaches. For robustness, we also report propensity score matching estimates in the appendix.

6.1. Identification 1: Cross-Group Comparisons

Let Y_{iw} denote the outcome variable (revenue, yield, tickets sold, or market share) of all itineraries belonging to OD i in Week w. Let τ_{iw} be the treatment status: τ_{iw} equals one if the itineraries within this OD are treated in Week w (and zero otherwise). Our goal is to estimate the treatment effect:

$$TE_{iw} = \mathbb{E}[Y_{iw} \mid \tau_{iw} = 1] - \mathbb{E}[Y_{iw} \mid \tau_{iw} = 0].$$

To retrieve this effect, we make four comparisons:

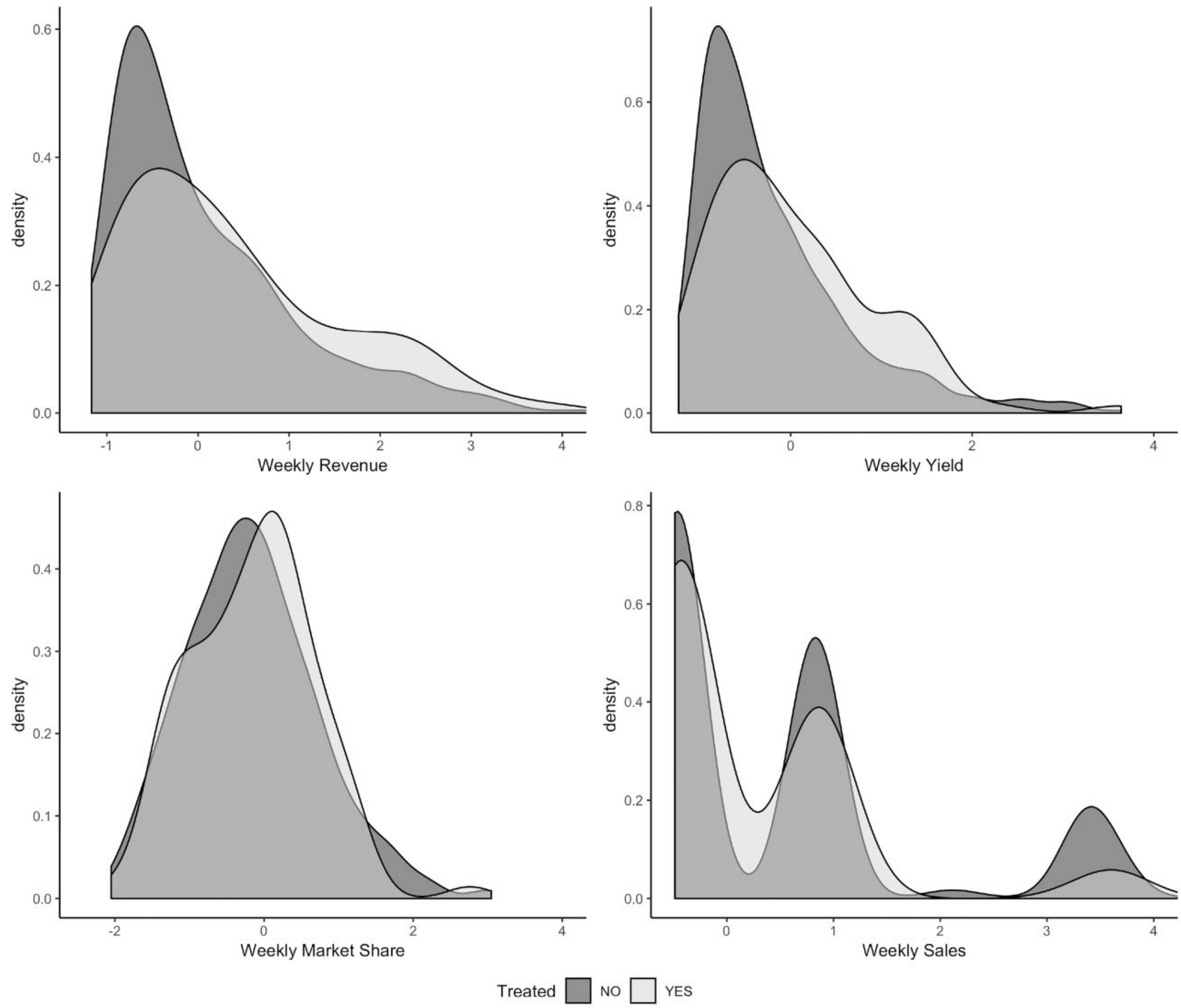
- *Test set 1 versus test set 2*. We compare each outcome variable of the treated and control observations. Recall that observations in TS_1 are treated during Weeks 1–4, whereas observations in TS_2 are treated during Weeks 5–8. The average treatment effect is thus equal to $\mathbb{E}[Y_{iw} - Y_{jw}]$, where $i \in TS_1, j \in TS_2$ in Weeks 1–4 and $i \in TS_2, j \in TS_1$ in Weeks 5–8.
- *Test set versus twin*. We compare each outcome variable of the treated observations to those in the set of twin ODs. The average treatment effect is thus equal to $\mathbb{E}[Y_{iw} - Y_{jw}]$, where $i \in TS_1, j \in Twin_1$ in Weeks 1–4 and $i \in TS_2, j \in Twin_2$ in Weeks 5–8.
- *Test set versus reverse*. We compare each outcome variable of the treated observations to those in the set of reverse ODs. The average treatment effect is thus equal to $\mathbb{E}[Y_{iw} - Y_{jw}]$, where $i \in TS_1, j \in Reverse_1$ in Weeks 1–4 and $i \in TS_2, j \in Reverse_2$ in Weeks 5–8.
- *Treated set versus all controls*. We compare each outcome variable of the treated observations to all nontreated observations. The average treatment effect is thus equal to $\mathbb{E}[Y_{iw} - Y_{jw}]$, where $i \in TS_1, j \notin TS_1$ in Weeks 1–4 and $i \in TS_2, j \notin TS_2$ in Weeks 5–8.

We formulate a linear regression model to make these comparisons:

$$Y_{iw} = Group_i + Week_w + \gamma Controls_{iw} + \delta\tau_{iw} + \varepsilon_{iw}, \quad (1)$$

where $Group_i$ captures the fixed effect of OD i's group (TS_1, TS_2, $Twin_1$, $Twin_2$, $Reverse_1$, $Reverse_2$), $Week_w$ captures week-level fixed effects, and $Controls_{iw}$ is a vector of idiosyncratic characteristics. This vector includes the eight variables defined in Section 5. It also includes controls for the itinerary's booking surcharges and a set of dummies that capture the proportion of bookings from each geographic location and purchase channel (e.g., the airline's website or a travel agency). Before adding these controls, we ran a variance inflation factor (VIF) test to rule out potential collinearity issues. In this specification, the average treatment effect is captured by δ.

Figure 6. Probability Density Function of Outcome Variables for the Treated and Control Observations



Note. The metrics on the *x* axes are measured in standard deviations, relative to the historical deseasonalized average.

6.2. Identification 2: Staggered Difference-in-Differences

We run a difference-in-differences (diff-in-diff) estimator to benchmark the performance of the treated groups, pre- and posttreatment, against the performance of the nontreated groups, pre- and posttreatment. Our analysis, however, differs from a standard diff-in-diff model in two ways. First, we have two treated groups that receive the treatment at different times (in typical models, we only have one treated group). Second, the treatment disappears after four weeks in each treated group (in typical models, the treatment runs until the end of the timespan).

Far from hampering identification, these two characteristics help us design a fine-grained diff-in-diff model to establish causality. Specifically, our approach follows Goodman-Bacon (2018), who derives a two-way diff-in-diff estimator that is applicable in contexts where there is variation in the treatment timing across groups. We can think of this setting as one with three groups and four treatments. The three groups are TS_1, TS_2, and a control group (i.e., the itineraries that were never treated). The four treatments happen

- In Week 1, when TS_1 receives a positive shock (moving from nontreated to treated);
- In Week 5, when TS_2 receives a positive shock (moving from nontreated to treated);
- In Week 5, when TS_1 receives a negative shock (moving from treated to nontreated); and
- In Week 9, when TS_2 receives a negative shock (moving from treated to nontreated).

Let g index the group, τ^{+}_{wg} be a dummy indicating if Group g received a positive shock in Week w, and τ^{-}_{wg}

be a dummy indicating if Group g received a negative shock in Week w. Let $Post^w$ be a dummy indicating if observations are recorded after Week w. We specify our diff-in-diff model:

$$\begin{aligned} Y_{igw} = {} & Group_g + Week_w + \gamma Controls_i \\ & + \sum_{w \in \{1,5,9\}} \theta_w Post^w + \sum_{w \in \{1,5\}} \psi_w^+ \tau_{wg}^+ \\ & + \sum_{w \in \{5,9\}} \psi_w^- \tau_{wg}^- + \sum_{w \in \{1,5\}} \beta_w^+ Post^w \cdot \tau_{wg}^+ \\ & + \sum_{w \in \{5,9\}} \beta_w^- Post^w \cdot \tau_{wg}^- + \varepsilon_{igw}. \end{aligned} \tag{2}$$

In line with Goodman-Bacon (2018), our specification includes group-level fixed effects ($Group_g$), time-period fixed effects ($Week_w$), and a vector of controls ($Controls_i$). The model-free nature of the diff-in-diff estimator lets us obtain consistent estimates without these controls—but including this vector improves the estimator's efficiency (Angrist and Pischke 2008, Lechner 2011). For robustness, we obtain estimates with and without controls. We also use robust standard errors and cluster all errors at the OD level, in line with Angrist and Pischke (2008, p. 237), who argue that "the simplest and most widely applied approach [to avoid a serial correlation issue] is simply to pass the clustering buck one level higher." With our specification, we pass the clustering buck from the lowest unit of analysis (i.e., the itinerary level) to the OD level. The number of clusters in our sample is above the minimum recommended threshold (Cameron and Miller 2015).

In this specification, the diff-in-diff coefficients are $\beta_1^+, \beta_5^+, \beta_5^-$, and β_9^-. This equation combines four diff-in-diff models. Therefore, we can make a sturdy causal argument if β_1^+ and β_5^+ are positive and statistically significant and β_5^- and β_9^- are negative and statistically significant.

7. Results

We prove our hypothesis in three parts. First, we establish the *baseline result*—we show, in Section 7.1, that under the current ladder-design practice—of matching airfare ladders—there are inefficiencies in the price path, particularly for high-quality itineraries. We also show that such inefficiencies can be mitigated by introducing an early price differential—for instance, by marking up the fare ladder.

Second, we establish the *mechanism*—we show, in Section 7.2, that the performance improvement is driven by a smoother price path. And third, we establish *boundaries* for our result.

7.1. Baseline Result

Table 5 shows a positive treatment effect for revenue. The effect is statistically significant. The effect is significant (at the 95% level) across all specifications, ranging between 0.21 and 0.34 standard deviation (where 0.21 and 0.34 correspond to the smallest and largest coefficient estimates). Our estimates also suggest that the airline's weekly yield increased by 0.21–0.25 standard deviations. In terms of tickets sold and market share, the effects are small in magnitude and lack statistical significance. Simply put, our treatment improves revenue and yield without sacrificing sales and market share.

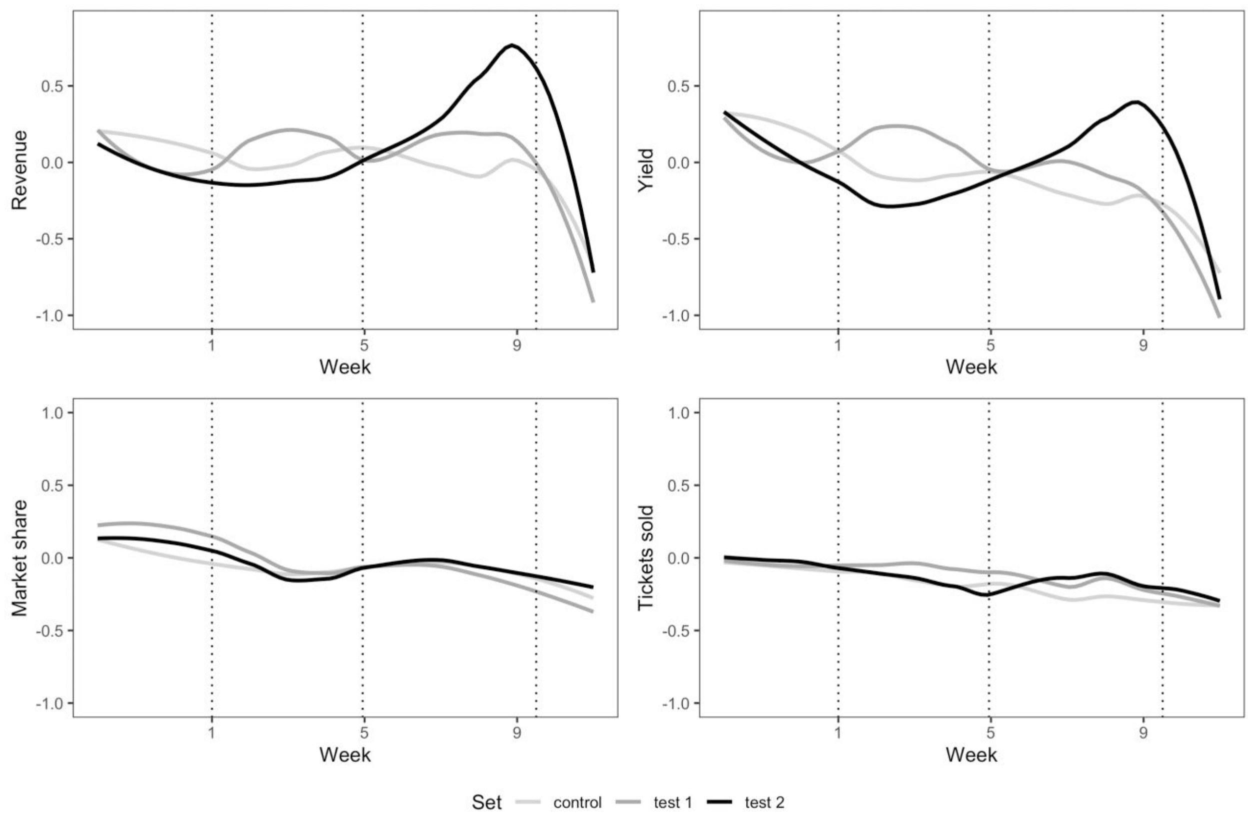
In the diff-in-diff specification, we obtain striking evidence of a treatment effect. Figure 7 shows that TS_1 and TS_2 increase revenue and yield during their treatment phase (Weeks 1–4 for TS_1 and Weeks 5–8 for TS_2), relative to other periods. The effect on market share and sales is visually insignificant.

Table 5. Estimates of Cross-Group Comparisons

Dependent variable	I	II	III	IV	V	VI	VII	VIII
Revenue	0.21**	0.29***	0.21***	0.34**	0.21**	0.26***	0.21**	0.34***
	(0.10)	(0.08)	(0.06)	(0.14)	(0.09)	(0.07)	(0.09)	(0.06)
Adjusted R^2	0.08	0.81	0.03	**0.88**	0.07	0.81	0.02	0.882
Yield	0.23***	0.24***	0.23***	0.21***	0.23***	0.25***	0.23***	0.22***
	(0.08)	(0.09)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)
Adjusted R^2	0.11	0.50	0.03	0.68	0.09	0.53	0.02	0.67
Tickets Sold	-0.04	−0.01	−0.28	−0.10	0.04	−0.11	0.04	−0.06
	(0.25)	(0.01)	(0.18)	(0.11)	(0.18)	(0.17)	(0.19)	(0.12)
Adjusted R^2	0.01	0.99	0.01	0.80	0.01	0.65	0.01	0.68
Market share	0.05	0.14	−0.097	−0.09	−0.15*	0.09	−0.15	−0.37***
	(0.11)	(0.10)	(0.10)	(0.10)	(0.088)	(0.10)	(0.09)	(0.10)
Adjusted R^2	0.00	0.40	0.00	0.80	0.01	0.38	0.00	0.31
Method	Test 1 versus test 2		Test versus twin		Test versus reverse		Test versus all	
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Itineraries	4,913	4,913	13,720	13,720	99,48	9,948	18,755	18,755

Note. The table reports the estimated coefficient values and standard errors (in parentheses) of the estimates using Equation (1).
*, **, and ***Significance levels at 10%, 5%, and 1%, respectively.

Figure 7. Time Series of the Outcome Variables



Notes. The x axes span from Week −4 to Week 12. The three dotted lines represent the shocks in Weeks 1, 5, and 9. This figure depicts time series of the treated and control observations using a locally weighted smoothing (LOESS) fit.

The numerical results—reported in Table 6—support the visual cues. This table shows results for three specifications for each outcome variable. The specifications in Columns I, IV, and VII truncate the sample at the end of Week 4, pretending the experiment ends at that point. This model is akin to a standard diff-in-diff model, with only a single treated group (TS_1) and one shock (β_1^+). In the specifications in Columns II, V, and VIII, we truncate the sample at the end of Week 8, assuming that the experiment ends at that point. This specification includes two treated groups and three shocks (β_1^+, β_5^+, and β_5^-). In Columns III, VI, and IX, we consider the entire timespan. This specification includes the three treated groups and all four shocks. The results show that, across *all* specifications, β_1^+ and β_5^+ are positive (and statistically significant) for both revenue and yield, whereas β_5^- and β_9^- are negative (and statistically significant) for both revenue and yield. This means that (i) a transition from a matching to a markup strategy significantly *increases* revenue and yield (by about 0.25 standard deviation); conversely, (ii) a transition from a markup to a matching strategy significantly *decreases* revenue and yield (by about 0.2 standard deviation). In contrast, the coefficients for market share and sales are statistically insignificant.

Our two models cross-validate each other. In particular, both models show that under current practice, the fare management systems lead to an inefficient price path (especially when it comes to high-quality itineraries) and that such inefficiencies can be mitigated by marking up the ladder—that is, by implementing quality-based fare management. In terms of magnitude, the gains are not trivial either—the airline manages to increase revenue and yield by 0.25 standard deviation. Furthermore, these benefits do not come at the cost of a decrease in total sales or market share.

7.2. Mechanism: The Price Path Dynamics

Section 7.1 establishes the superiority of one ladder-design policy over the other, given the current revenue management algorithms. When explaining the mechanism driving this revenue boost, our intuition is to deduce that lead-in matching underprices high-quality itineraries; that is, customers are willing to pay more than the fares set under a matched ladder. However, this conjecture is implausible. If the itineraries were indeed underpriced, the revenue management algorithms would have moved the fares up the ladder and

Table 6. Estimates of Difference-in-Differences

	Revenue			*Yield*			*Tickets Sold*			*Market share*		
Dependent variable	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
TS_1	0.16***	0.16***	0.17***	-0.06***	0.17***	0.17***	−0.06	-0.078*	−0.07	−1.15***	−1.1***	−1.16***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
TS_2	0.38***	0.38***	0.39***		0.38***	0.39***	−0.03	−0.04	−0.03	−0.95***	−0.97***	−0.98***
	(0.02)	(0.02)	(0.02)		(0.02)	(0.02)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
Post 1	0.21***	0.21***	0.26***	0.05	0.06	0.08	0.47***	0.84***	0.85***	−1.03***	−1.03***	−0.93***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.13)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)
Post 5		-0.04	0.07		-0.09	−0.11*		-0.11	0.51***		0.06	−0.04
		(0.05)	(0.05)		(0.06)	(0.06)		(0.13)	(0.13)		(0.12)	(0.12)
Post 9			-0.33***			−0.10			0.31**			−0.16
			(0.06)			(0.07)			(0.16)			(0.12)
β_1^+	0.21***	0.21***	0.21***	0.23***	0.22***	0.22***	-0.10	-0.10	-0.10	-0.13	-0.13	−0.13
	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.14)	(0.15)	(0.15)	(0.13)	(0.13)	(0.13)
β_5^-		−0.19**	−0.20***		−0.20**	−0.21**		0.31	0.18		-0.06	−0.10
		(0.08)	(0.08)		(0.09)	(0.09)		(0.20)	(0.19)		(0.17)	(0.16)
β_5^+		0.23***	0.23***		0.230***	0.23***		−0.57*	−0.58*		0.03	0.02
		(0.06)	(0.06)		(0.07)	(0.07)		(0.44)	(0.45)		(0.12)	(0.12)
β_9^-			−0.24**			−0.24*			0.14			0.00
			(0.12)			(0.14)			(0.31)			(0.26)
Itineraries	12,590	13,119	13,280	*12,590*	13,119	13,280	12,590	13,119	13,280	11,344	11,827	11,971
Adjusted R^2	0.08	0.08	0.08	0.07	0.07	0.07	0.01	0.01	0.01	0.04	0.04	0.04

Notes. The specifications in Columns I, IV, and VII truncate the sample at the end of Week 4. The specifications in Columns II, V, and VIII truncate the sample at the end of Week 8. The specifications in Columns III, VI, and IX do not truncate the sample. The table reports the estimated coefficient values and standard errors (in parentheses) of the diff-in-diff estimates in Equation (2).
*, **, and ***Significance levels at 10%, 5%, and 1%, respectively.

captured this surplus since these algorithms are precisely designed to trim mispricing.

In Section 3, we theorized a more plausible mechanism: marking up the ladder boosts revenue, we argued, through a smoother price path. By marking up the ladder, the airline substitutes early sales (at a low fare) in exchange for late sales (at a higher fare). This is because, by keeping observed fares higher earlier in the selling season, the price path will generate smaller fare increases earlier in the selling season and more revenues later in the selling season.

The experimental results support our hypothesis. We arrive at this finding by analyzing the itineraries day by day, in Figure 8. The leftmost plot compares treated itineraries with control- roup itineraries (that have the same departure and arrival dates) across the 28-day experimental trial. If the bar is positive on a given day, the treated itineraries outperformed the control group itineraries (in terms of revenue or sales). However, if the bar is negative, the control group itineraries performed better. The rightmost plot reports the fare percentage changes over time.

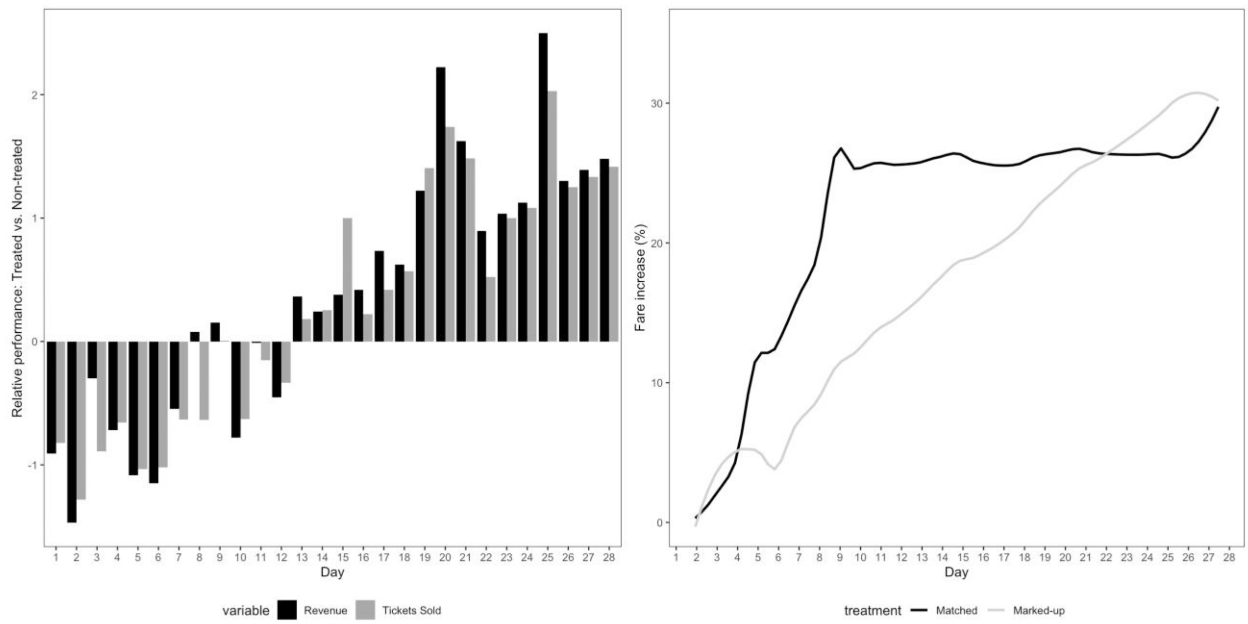
These plots show that the matched strategy had an enthusiastic start but a lackluster closing. Specifically, itineraries with a matched strategy sold significantly more seats at the outset. Accordingly, the revenue management systems moved up the ladder quickly, causing a drastic fare spike. However, after this initial price hike, sales slowed down and the fares stabilized. Itineraries with a markup strategy, in contrast, underperformed at the beginning of the selling season but outperformed later on. We find that the sales patterns are steadier under the markup strategy, leading to a smoother fare increase through the selling season.

To understand the driver behind this dynamic, consider a market with two airlines competing with the same ladder. The airline with the higher-quality itinerary will provide a "bargain" early in the season, so consumers will rush to buy this itinerary. Accordingly, the airline will get a large inflow of revenue and fill most seats early on. The revenue management systems will then move the fare up the ladder (to sell the few remaining seats at a higher price). Thus, the airline will have a low stream of sales later in the selling season. Simply put, the revenue management algorithms will give rise to a steep price path in the early selling season that then will flatten in the late part of the season.[3]

Let us now consider the case where the high-quality itinerary has a marked-up ladder. In this case, the itinerary will sell fewer tickets early on, precisely because it no longer offers a "bargain." The airline will thus have modest sales in the early season, which are accompanied by slower price increases. The airline will, however, capture a larger number of sales later on because of the lower price tag (relative to the matched strategy). The price path will thus evolve in a smoother fashion throughout the selling season.

A caveat: To fully document this mechanism, we would need to track sales across the entire selling season, through a year-long experiment. Our experiment only

Figure 8. Price Path: Treated and Control Itineraries



Notes. (Left) The y axis displays the revenue and sales of treated itineraries, relative to control-group itineraries, in terms of standard deviations. If the bar is positive, treated itineraries outperformed the control-group itineraries. The x axis reports the days since beginning of the treatment. (Right) The figure shows the fare percentage increase of treated and control itineraries. The x axis reports the days since the beginning of the treatment.

lasted eight weeks. However, the "hot part" of the selling season, for an average itinerary, spans four to seven weeks. Specifically, the majority of the sales bulk up in a handful of weeks, just before the departure date. Also, remember that we chose to experiment with itineraries that were about to enter this hot part of the selling season (by selecting itineraries that depart within 120 days). In this sense, we capture the most important segment of the season and, thus, this limitation is not too stringent.

7.3. Boundary Conditions

Thus far, we focused on the average treatment effect under both strategies: lead-in matching and quality-based fare management. However, our stylized model shows that it would be an oversimplification to assume that lead-in fare matching is always advantageous. For example, our results show that lead-in markups have a tampered advantage—or even a counterproductive one—when the itineraries attract a quality-insensitive customer base or when demand is weak altogether. This makes sense: if consumers do not care about quality enough, they will be less prone to accepting a markup, and the mechanism behind our hypothesis will be feeble.

In our data, we can find variables that reliably proxy for demand and quality sensitivity. In particular, we can measure the strength of demand by studying (i) the airport traffic at the origin and the destination airports, (ii) the departure time (midday flights face significantly lower demand than early-morning and late-afternoon flights), and (iii) the number of competitors. We can also proxy for quality sensitivity by looking at (i) the airline's relative market share (linked to brand loyalty), and (ii) the length of the trip (customers are more sensitive to quality differences on longer flights).

To examine the boundaries of our results, we investigate how the two pricing strategies perform as a function of these itinerary characteristics. Specifically, we estimate heterogeneous treatment effects to determine the factors that drive customers' quality sensitivity and market demand and thus make lead-in markups relatively more (or less) effective than lead-in matching.

To estimate these effects, researchers have traditionally used nearest-neighbor matching and kernel methods. These methods have asymptotic shortcomings, especially when estimating several moderators simultaneously. To address this issue, Wager and Athey (2018) proposed a *causal forest* method that uses random forest algorithms for causal inference. The authors demonstrate, via simulations, that causal forest methods outperform k-nearest-neighbor matching in terms of bias and variance.[4]

We apply the causal forest method in our sample by considering the itinerary characteristics from Table 3. An advantage of these moderators is that they are easily interpretable and quantifiable, meaning that airlines can use them to determine whether to adopt lead-in matching or lead-in markups on a particular itinerary.

Figure 9 plots the predicted treatment effect on revenue, yield, tickets sold, and market share as a function of each moderator. Table 7 provides a linear fit of the prediction. Our results show that, all else being equal, the quality-based fare management strategy is most effective when

1. The destination has high traffic
2. The OD is served by a larger number of airlines
3. The departure is in the early morning or late at night (i.e., during high-demand times)
4. The trip length is longer
5. The airline has a high degree of market share

Collectively, the first three results show that the treatment has a larger effect in market segments where demand is strong. In terms of points 4 and 5, the causal forest predictions show that the treatment effect is strongest in market segments with more quality-sensitive consumers. Let us explain in detail:

With regards to Result (1), itineraries with high-traffic destinations have a larger mass of consumers and typically sell out fast. Thus, the airline benefits by factoring the quality differential at the outset—that is, in the design of the ladder—given that it will be able to capture enough quality-sensitive customers. Moreover, competing airlines are likely to close lower inventory classes faster, so the airline with a marked-up ladder becomes more competitive earlier in the season. This result is in line with the model results, which state that a matching strategy is more effective when demand is strong.

With regards to Result (2), the correlation between the number of competitors and the demand for the OD is 0.88, meaning that the supplier base serves a strong proxy for the number of potential customers or, put another way, the strength of demand.

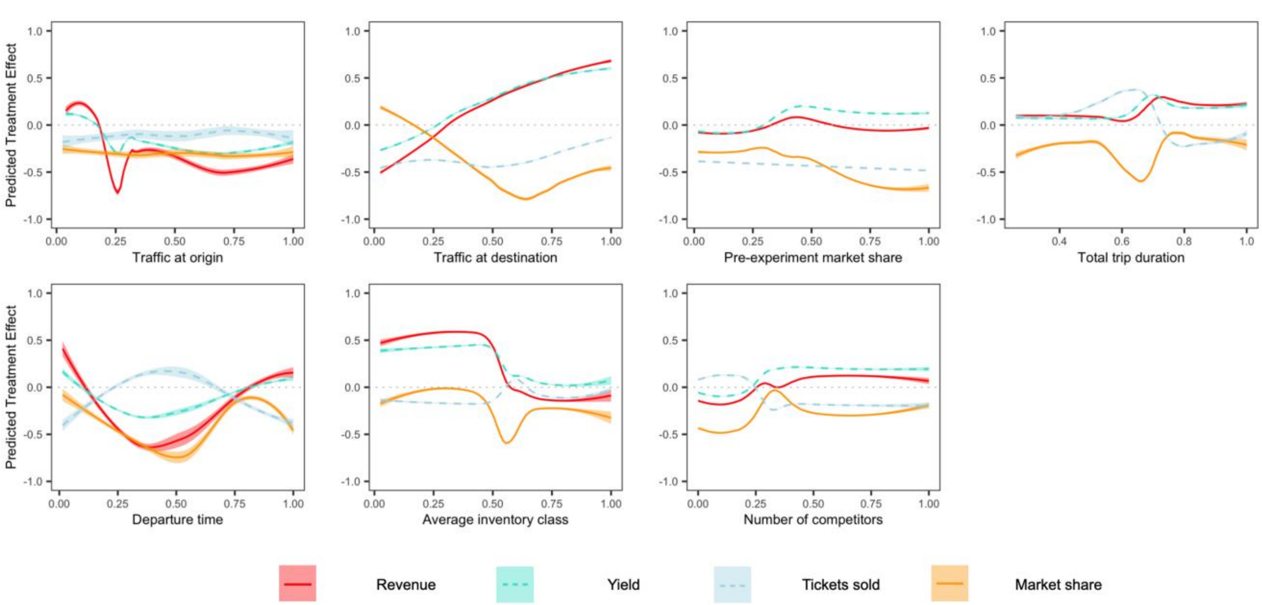
In terms of Result (3), morning and evening itineraries are attractive to most travelers—that is, the demand for flights is U-shaped within a daily schedule. In this fashion, proactive ladder differentiation becomes a better strategy. The plots suggest a sharp convex pattern, in revenue and yield, for departure time. Thus, we ran an additional regression specification by including a quadratic term for *Departure time*. The results confirm the plot's pattern: The linear and quadratic coefficients are equal to (i) -0.14 and 3.46 for *Revenue* and (ii) -0.07 and 1.74 for *Yield*.

With regards to Result (4), segments with low market share cater mainly to loyal customers (in fact, the correlation between registered loyalty customers and market share is -0.78). Because loyal customers are less price sensitive, lead-in markups are more effective by factoring quality into the ladder.

With regards to Result (5), customers are willing to pay a large premium for a high-quality itinerary on a long trip. Accordingly, more of them are quality sensitive, making it profitable to introduce quality differentiation ex ante (i.e., in the ladder design).

As explained previously, these results show that itineraries with characteristics that attract quality-sensitive travelers—longer trip times, more competitors, or a large demand share—perform better under a markup strategy. The results also show that, as hinted by our model, the treatment effect is stronger when there is a sufficiently high demand base—high destination traffic, midday departures. These results contribute to the empirical airline management literature: they confirm the impact

Figure 9. (Color online) LOESS Fit for Heterogeneous Treatment Effects



Notes. The figure shows the predicted treatment effect, using the causal forest method from Wager and Athey (2018). For interpretability, the independent variables (except *Number of competitors)* have been normalized to represent the quantiles of the distribution.

Table 7. Linear Fit of Causal Forest Predictions

Dependent variable	Revenue	Yield	Tickets sold	Market share
Traffic at origin	0.002	−0.0001	0.0004*	−0.001*
	(0.013)	(0.009)	(0.0001)	(0.0003)
Traffic at destination	0.034***	0.092***	−0.0104*	0.002***
	(0.010)	(0.006)	(0.00001)	(0.0002)
Pre-experiment market share	−0.304***	−0.208***	0.0051**	0.004***
	(0.012)	(0.008)	(0.00002)	(0.0003)
Trip duration	0.168***	0.076***	0.0061	−0.008***
	(0.018)	(0.012)	(0.0001)	(0.0004)
Departure time	0.052***	0.004	0.0061	0.001***
	(0.010)	(0.006)	(0.0001)	(0.0002)
Average inventory class	−0.618***	−0.430***	0.0021	−0.002***
	(0.024)	(0.015)	(0.0032)	(0.001)
Number of competitors	0.005	0.026**	0.0063	0.001**
	(0.019)	(0.012)	(0.0002)	(0.0004)
OD weeks	789	789	789	789
Adjusted R^2	0.673	0.727	0.698	0.527

*, **, and ***Significance levels at 10%, 5%, and 1%, respectively.

of trip duration and departure time on passengers' willingness to pay, and identify the effects of several new moderators (e.g., *Traffic at destination*, *Market share*, *Average inventory class*, and *Number of competitors*).

We can use these results to provide actionable insights that airline managers can use to decide between the two pricing strategies: lead-in matching and quality-based fare management. Suppose, for example, that our airline partner creates a decision framework where it *only* implements our quality-based policy across itineraries where the treatment effect is positive; for the rest, the airline continues to apply the lead-in matching strategy. Our results predict that this strategy will increase revenue and yield on these itineraries by 0.82 and 1.24 standard deviations on average.

7.4. Robustness Analysis

We perform a series of robustness checks, which are presented in Appendix B. We begin in Sections B.1 and B.2, where we perform cross-sectional and intertemporal placebo tests, respectively, to gauge the robustness of our results to temporal and cross sectional autocorrelation in the error terms. Then, in Appendix B.3, we present a spillovers-robust diff-in-diff model to account for network spillovers stemming from our experiment. Finally, in Appendix B.4, we establish the robustness of the results to the time until the flight's departure. Despite minor numerical changes in the estimates, the main results of our model are robust to these alternative specifications; see the appendix for more details.

8. Discussion of Results

8.1. Implications for Airfare Pricing and Revenue Management Systems

In an ideal world, airlines would determine airfares by optimizing prices within a continuous space for a given booking request (Gallego and Van Ryzin 1994, Bitran and Caldentey 2003, Talluri et al. 2004). Presently, airlines find it impossible to implement airfare pricing without relying on fare classes. The impossibility stems out of two sources. First, optimizing a fare ladder via continuous pricing poses severe computational challenges, which require dedicated approximate dynamic programming algorithms that are beyond our reach. Second, airlines still rely on pricing software solutions that can only accommodate a coarse set of price points. Such hurdles have led the industry to embrace a two-step pricing system to control the evolution of the price path. This two-step approach consists of (i) designing a fare ladder ex ante and (ii) implementing revenue management algorithms ex post, which act as capacity controls or bid price controls throughout the fare ladder.

Currently, the operationalization of this two-step procedure goes hand-in-hand with two practices: ladder matching, on the one hand, and ex post differentiation, on the other. The idea is that ladder matching allows competitors to simplify the design of the price path so that, in turn, the revenue management algorithms can differentiate the displayed fares between competing airlines.

Our experiment shows that the joint application of these two practices leaves unrealized revenue on the table. This result offers two interpretations, leading to two possible implementations.

- *Interpretation 1 (ladder design inefficiencies).* The first interpretation is that ladder matching leads to a suboptimal price path (i.e., a suboptimal fare ladder). This interpretation suggests that airlines should revise the architecture of the airfare ladder and attempt to implement quality differences into its design instead of systematically relying on matching strategies.
- *Interpretation 2 (revenue management inefficiencies).* The second interpretation is that existing revenue

management systems are unavailing under a matched ladder strategy. This interpretation suggests that airlines should revise their dynamic pricing algorithms to implement quality-based fare management. In a bid price control mechanism, this implementation can set up higher bid prices on high-quality itineraries. In a quantity-based control mechanism, this implementation can further restrict the number of seats available at low prices during the early selling season. Either way, this implementation would smoothen out the price path across the selling season and increase overall revenues.

These two implementations come with tradeoffs. On the one hand, changes in the design of fare ladders are relatively easy to implement and, as our results show, yield nontrivial revenue gains. However, competitors can easily observe the changes across fare ladders, so they may react to this change. As such, redesigning the ladder might not be as beneficial in the long term. Changes in the revenue management algorithms are more time-consuming, more complex and more expensive to implement. On the other hand, competitors would find it harder to observe and mimic the changes, which could lead to a more sustained edge in the long term. An airline would need to balance such trade-offs when deciding which implementation strategy to pursue.

These results come at a time of gradual transition in airline distribution capabilities toward continuous pricing, to enable more granular fare ladders and more pricing flexibility within each class. This transition will require extensive research to design more sophisticated and more complex revenue management algorithms. In this regard, our findings provide interpretable guidelines to introduce quality differentiation into the dynamic optimization of price paths. Namely, next-generation revenue management systems should mark up airfares early in the selling season on high-quality itineraries to induce a smoother price path and a dynamic substitution in sales throughout the selling season. In addition, our experiment identifies ODs where existing pricing and revenue management systems are effective and, at the same time, ODs where new pricing and revenue management algorithms are most needed—therefore guiding airlines' investments into new pricing solutions. Ultimately, airline managers should take our results as a call to improve the joint coordination of the fare ladder and revenue management algorithms to maximize revenues.

8.2. Limitations

Our experiment comes with five limitations. The first limitation is that we ran our field experiment for two months (instead of one or two years), meaning that we were unable to record the long-term dynamics of our experiment.

This first limitation gives rise to the second limitation, namely, that we were unable to observe how competitors react to fare-ladder increases during this short period. If a ladder markup strategy benefits one airline, all airlines would profit by deviating from the matching baseline on their own high-quality itineraries. At the same time, we could expect a reaction from airlines with low-quality itineraries.

The third limitation is that the experiment was conducted only across international markets. That said, the dynamics of pricing and revenue management are remarkably similar between domestic and international itineraries. Our results would likely have a similar impact on domestic markets as they did on international markets in our experiment.

The fourth limitation is that the experiment was deployed exclusively across high-quality itineraries, namely, itineraries with a QSI above 0.5. Accordingly, we cannot make conclusive claims about the differential impact of QSI on the magnitude of revenue increase.

A fifth limitation is that the experiment was run on a small set of 22 O-D markets, with limitations on traffic, volume, and timing. As argued earlier, this set of markets still involved a complex implementation that impacted nearly 5,000 itineraries and 100,000 bookings. Yet, this limitation potentially restricts the generalizability of our findings to other ODs where the airline offers high-quality itineraries.

8.3. Moving Forward

The intent of this paper is to introduce a much-needed discussion on the way airlines design their fare management systems—especially in how the design of a ladder affects the optimality of the revenue management algorithms. This is because, despite the large inflow of research in airfare dynamic pricing, the design of the airfare has been mainly taken as a given input. Future research can build upon our study by, for example, analyzing how to fine tune the design of the ladder. Our experiment applies a uniform markup, as opposed to tuning the markup at the OD level or at the itinerary level. But researchers could follow up by conducting more granular experiments to find the optimal markup based on the characteristics of ODs and itineraries.

Future research could also examine how to optimize the revenue management algorithms in a differentiated ladder. If airlines move toward quality-differentiated ladders, they might also want to upgrade their revenue management systems. Thus far, these systems have been optimized under the presumption that the market has a matched ladder.

A final overarching question is whether the market, as a whole, becomes more or less efficient when airlines implement quality-based fare management when designing their ladder or, alternatively, when designing their revenue management algorithms.

Acknowledgments

The author's sincerely thank our airline partner for the collaboration. The author's also thank the department editor (Vishal Gaur), the associate editor, and the three anonymous referees for their insightful comments, which have helped improve this paper.

Appendix A. Propensity Score Matching

We use propensity score matching (PSM) as an alternative identification strategy. To perform this analysis, we match every treated itinerary with nontreated itineraries that resemble as much as possible the treated one. Our matching strategy uses the continuous variables outlined in Figure 5. To obtain each observation's propensity score, we use a parametric generalized linear model that estimates approximate standard errors, with fixed weights, on the treatment effects.

After estimating the propensity score, we verify that the treatment and control groups have similar distributional properties across each matching variable. In line with the technique's *modus operandi*, we use propensity score reweighing to rebalance the distributions.

A.1. Distributional Balancing

PSM depends on the assumption that the treatment and control samples are identically distributed across all covariates (i.e., that the distributions are "balanced"). If the distributions are unbalanced—as it is often the case—we can adjust them by reweighing each observation (whereas it is acceptable to have unbalanced distributions at the outset, our analysis will fail if we cannot find a suitable weight vector to rebalance them). Specifically, we use a generalized boosted model to find a weight vector that optimizes the balance in the distributional means, variances, and cumulative distributions (Austin 2011, Austin and Stuart 2015):

1. *Balance of means.* To measure the balance of the distributions' first moments, we measure the standardized bias for each variable. We calculate the standardized bias of covariate x by measuring the absolute difference of the means, $|\mu_x^{treatment} - \mu_x^{control}|$, and dividing this difference by the pooled standard deviation. As a rule of thumb, the adjusted standardized differences should be smaller than 0.1 (Stuart et al. 2013). Figure A.1 and Table A.1 illustrate the standardized bias for each covariate, before and after reweighing the distributions. These two exhibits show that the distributional means are highly balanced, meaning that the reweighing process was successful.

2. *Balance of variances.* To explore the balance of the distributions' variances, we analyze the ratio of the treatment and control groups' variance. By convention, we place the largest variance in the numerator; a ratio of one means that the variances are perfectly balanced and, as a rule of thumb, a ratio below two is acceptable after adjusting the distributions (Rubin 2001). Table A.1 shows that the distributions' variances comfortably satisfy this requirement.

3. *Balance of cumulative distribution*. We explore the balance of the cumulative distribution functions with the Kolmogorov-Smirnov statistic, which measures the maximum distance between the support of the cumulative distribution functions. This statistic ranges from zero (perfect balance) to one (full imbalance). By convention, a value below 0.05 is recommended after adjusting. Table A.1 shows that all our adjusted covariates meet this recommendation.

A.2. Results

After balancing the distributions, we estimate the average treatment effect of the weighted distributions. For robustness, we also use nearest-neighbor (NN) matching, which estimates the average treatment effect after matching

Figure A.1. Standardized Bias for the Adjusted and Unadjusted Distributions

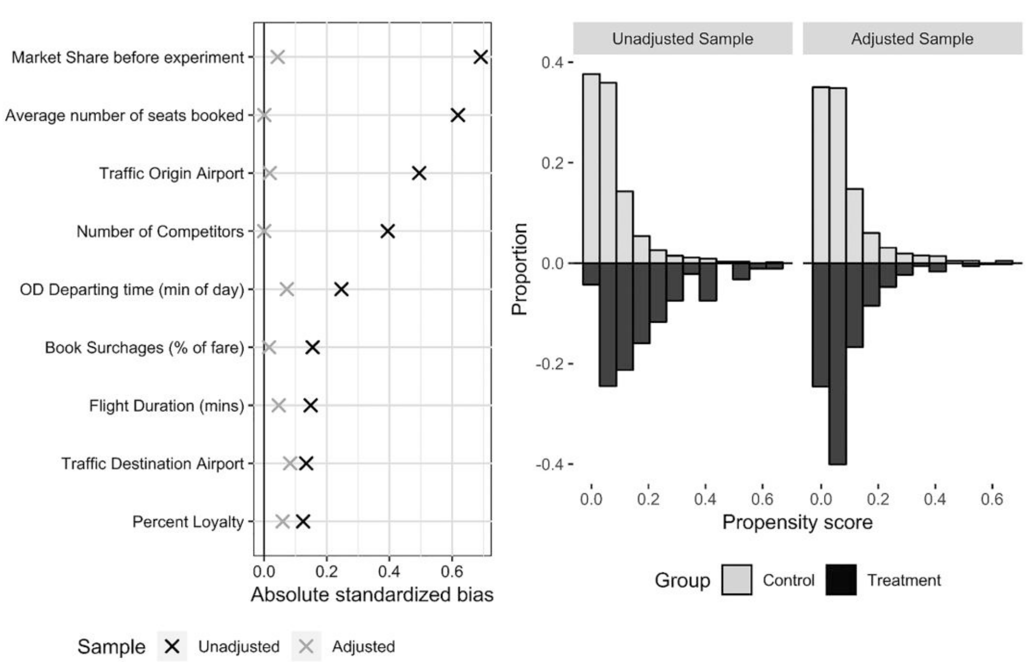


Table A.1. Statistics for the Unadjusted and Adjusted Distributions of the Raw Data

	Standardized bias		Variance ratio		Kolmogorov-Smirnov	
Covariate	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Percent loyalty	0.12	0.05	1.02	1.02	0.10	0.01
Book surcharges	0.16	0.03	1.22	1.53	0.13	0.02
Booking capacity	0.60	0.00	2.23	1.12	0.28	0.01
Competitor airlines	0.39	0.00	2.03	1.55	0.15	0.00
Pre-experiment market share	0.66	0.05	1.46	1.92	0.31	0.02
Traffic origin	0.49	0.04	2.34	1.19	0.22	0.01
Traffic destination	0.06	0.09	1.25	1.03	0.13	0.02
Departing time	0.26	0.07	1.24	1.04	0.23	0.01
Trip duration	0.15	0.05	1.48	1.55	0.21	0.03

the closest K neighbors of every treated observation (allowing for replacement); we run four estimates by letting K equal one, two, three, and four.

Our matching results (reported in Table A.2) confirm the same insights as our baseline results. In particular, the average treatment effect is positive and significant at the 5% level across all specifications for revenue and yield. The magnitude of the coefficients ranges from 0.33 to 0.43 for revenue, and from 0.19 to 0.42 for yield—which is also consistent with our baseline analysis. The estimates also show a statistically insignificant change in tickets sold and market share.

Appendix B. Robustness Analysis

B.1. Cross-Sectional Placebo Test

Our data set is large—with thousands of itineraries—but our cross-section of markets is small—with only 22 ODs. This means that we rely on temporal variation when calculating standard errors. In our specification, we permit general autocorrelation within time but no autocorrelation across time. Thus, persistent temporal shocks may bias our standard errors, leading to spurious results. This problem is common in field experiments. For example, Bray et al. (2016) studied judiciary task juggling but were only able to influence the behavior of six judges, and Stamatopoulos et al. (2017) tested the effects of electronic shelf labels in retail but could only study two treated stores.

To determine whether our results are artifacts of a small OD cross section, we conduct a placebo test. Specifically, we create 10,000 new samples by randomly assigning ODs to "treated" and "control" groups in each sample. We re-estimate our models across these 10,000 synthetic samples. Figure B.1 benchmarks the estimates of the 10,000 synthetic samples with the estimates of our true sample.

If our true estimates were artifacts of cross-correlation, they would not stand out relative to the placebo estimates. But Figure B.1 shows that our true estimates stand out in all cases. That is, our true estimates, which are normalized to zero, are always at the far end of the distribution. Furthermore, unlike our true estimates, the placebo estimates are statistically insignificant in 98.51% of the simulations. These results suggest that our findings are not artifacts of a small cross-section.

B.2. Intertemporal Placebo Test

Hypothetically, there could be something "special" about the treatment dates that artificially drove our results. For instance, the selected months could represent a favorable time period for the treated ODs. To rule out the impact of these confounders, we adopt a placebo test similar to the one adopted by Monfared and Pavlov (2017).

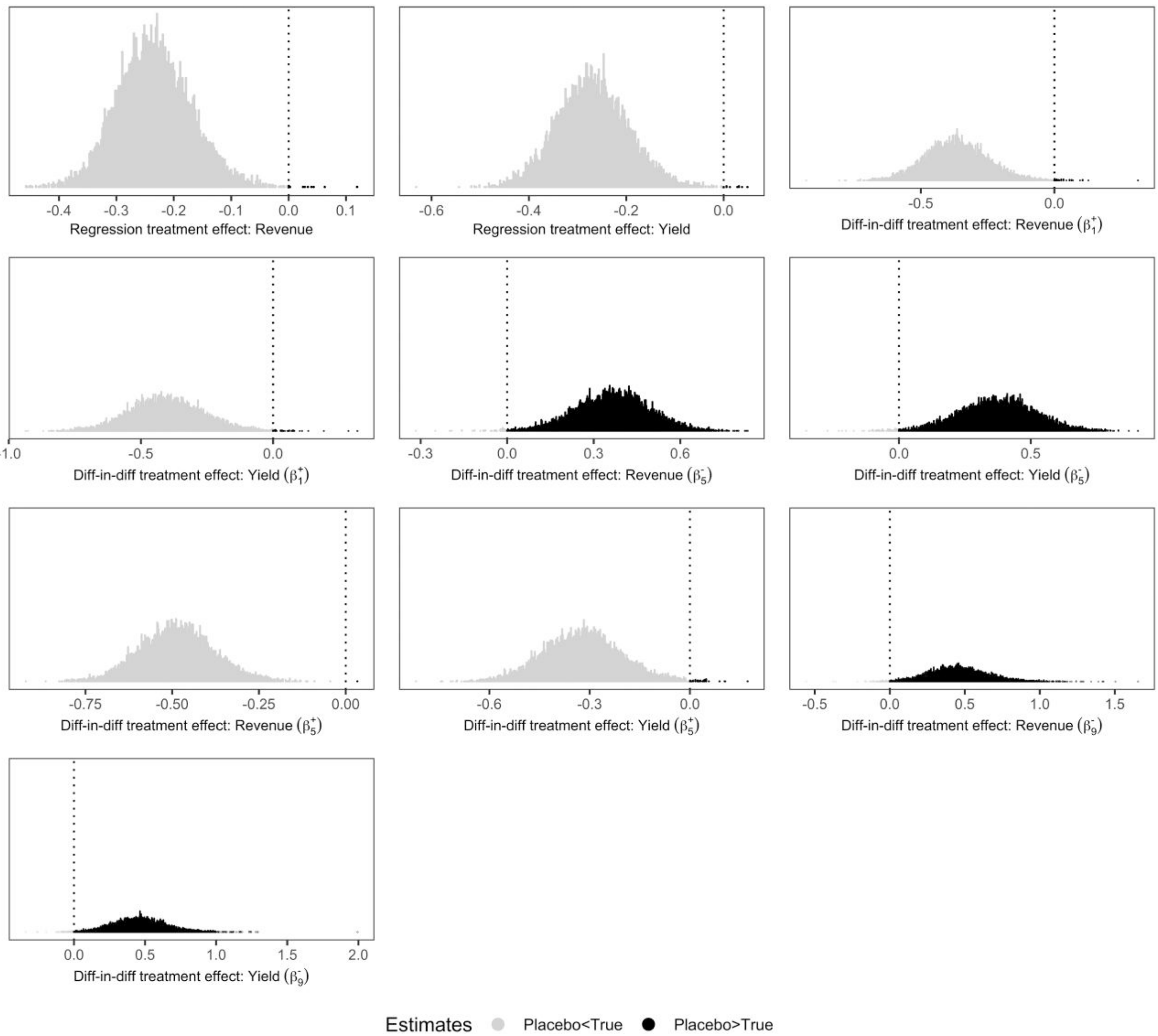
In this placebo test, we take our original data set and replicate it 100 times. In each case, we pretend that the treatment was applied during a different period from the "true" experiment, going back one week at a time for 100 weeks. Specifically, recall that our "true" treatment was applied from Week 1 to Week 4 on Group TS_1 and from Week 5 to Week 8 on Group TS_2. Thus, in the first synthetic data set, we pretend that the treatment was applied from Week −4 to Week −1 (to TS_1) and from Week 1 to Week 4 (to TS_2); in the second synthetic data set, we assume that the treatment was

Table A.2. Estimates of PSM

Dependent variable	I	II	III	IV	IV
Revenue	0.33***	0.43***	0.39**	0.39**	0.42***
	(0.06)	(0.18)	(0.19)	(0.20)	(0.18)
Yield	0.19***	0.42***	0.34**	0.35**	0.40***
	(0.05)	(0.15)	(0.16)	(0.16)	(0.15)
Tickets sold	−0.06	−0.04	−0.04	−0.04	−0.04
	(0.15)	(0.04)	(0.04)	(0.03)	(0.05)
Market share	−**0.**38	−0.27*	−0.24*	−0.21	−0.22*
	(0.50)	(0.134)	(0.14)	(0.14)	(0.13)
Matching method	PSM	NN: 1 neighbors	NN: 2 neighbors	NN:3 neighbors	NN:4 neighbors
OD weeks	13,118	13,118	13,118	13,118	13,118

Notes. The table reports the estimated coefficients and standard errors (in parentheses) for different matching criteria.
*, **, and ***Significance levels at 10%, 5%, and 1%, respectively.

Figure B.1. Cross-Sectional Placebo Test



Notes. This plot explores the robustness of our estimates at the cross-sectional level. We run equivalent regressions for 10,000 synthetic data sets. We construct the synthetic data sets by randomly selecting "treated" and "control" itineraries, while fixing the fraction of treated itineraries in the sample. We then plot the distribution of the synthetic estimates, normalized with respect to our true estimates (specifically, we plot the distribution of the difference between the synthetic and true estimates, divided by the pooled standard deviation). We color black the placebo estimates with more extreme results than our true estimates. Those with less extreme results are colored gray.

applied from Week -5 to Week -2 (to TS_1) and from Week -1 to Week 3 (to TS_2); and, more generally, on the nth synthetic data set—for n between 1 and 100—we assume that the treatment was applied from Week $-(n+3)$ to Week $-n$ (to TS_1) and from Week $-n+1$ to Week $-n+4$ (to TS_2). For expositional purposes, we do not have Week 0. All other aspects of the data set remain unchanged.

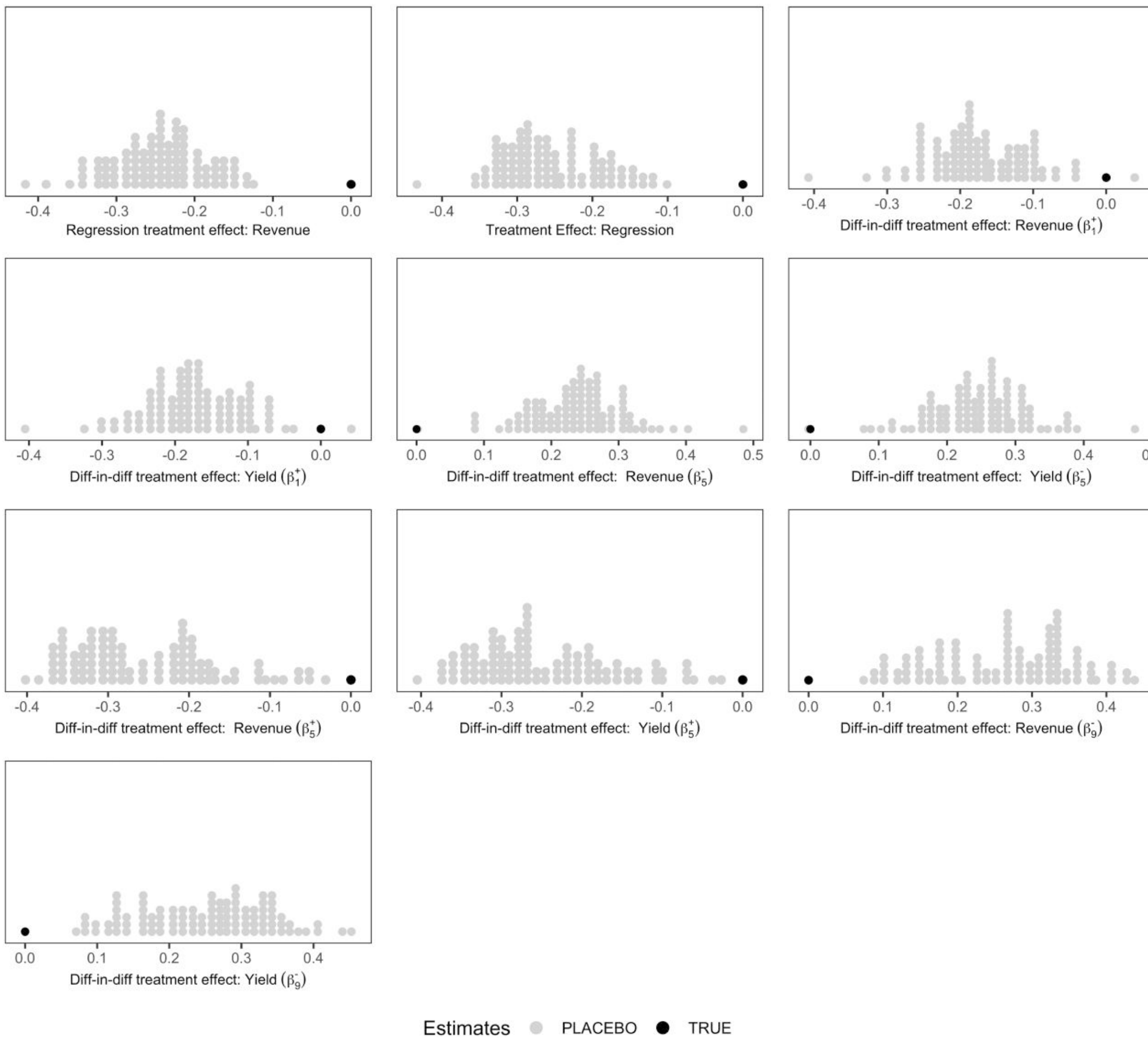
Figure B.2 reports the distribution of placebo estimates, which indicate that our true estimates stand out in all cases. Therefore, it is highly unlikely that our results are artifacts of time-correlated confounders.

B.3. Spillovers-Robust Diff-in-Diff Estimates

The validity of our estimates relies on the stable unit treatment value assumption (SUTVA) that, in our context, means that the treatment should not meaningfully spill over across itineraries in the control group. Simply put, untreated itineraries should not be affected by a change in price to the treated ODs. This assumption would be tenuous if a change in the fares of the treated group were to shift the demand (or prices) across other ODs in the network.

It is likely that a systematic change in the fares of the 22 ODs will have some effect across other itineraries in the airfare network. This type of spillover effect is not uniform, however, given that network spillovers are more prominent across some ODs. As a case in point, a change in the fare of the {Boston-Austin} OD is more likely to affect the {Boston-Houston} OD than the {Buenos Aires-Bogota} OD. Presumably, spillover effects would be more prominent across ODs that share flight legs with treated ODs, or those that have common

Figure B.2. Intertemporal Placebo Test



Notes. This plot explores the robustness of our estimates at the intertemporal level. We run equivalent regressions for 100 synthetic data sets. We construct the synthetic data sets by alternating the treated and control weeks, going back in time in one-week increments (starting from the actual treated date). We then plot the distribution of the synthetic estimates, normalized with respect to our true estimates (specifically, we plot the distribution of the difference between the synthetic and true estimates, divided by the pooled standard deviation). We color black the placebo estimates with more extreme results than our true estimates. Those with less extreme results are colored gray.

airport terminals, or a high demand correlation with the treated ODs.

To deal with this type of econometric issue, a growing number of papers have recently begun proposing structural solutions that are robust to network spillover effects in diff-in-diff models. Most notably, Angelucci and Di Maro (2016), Goodman-Bacon (2018), and Clarke (2019) propose a class of identification strategies that retrieves unbiased diff-in-diff estimates when a violation of the SUTVA arises—precisely, because of spillover effects from treated units to untreated units. These methods allow to measure (nonparametrically) the magnitude of the spillover effect across the network. Berg et al. (2020) and Butts (2022) generalize Clarke's method by relaxing the functional form of the error term.

We rely on the previous methods—primarily on the model of Butts (2022)—to design a diff-in-diff estimator that is robust to SUTVA violations (because of network spillover effects). Specifically, our model obtains diff-in-diff estimates that are robust to SUTVA violations (because of network spillovers) via the following specification:

$$\begin{aligned} Y_{igw} = {} & Group_t + Week_w + \gamma Controls_i + \sum_{w\in\{1,5,9\}} \theta_w Post^w \\ & + \sum_{w\in\{1,5,9\}} \psi_w^+ \tau^+ + \sum_{w\in\{1,5,9\}} \psi_w^- \tau^- \\ & + \underbrace{\sum_{w\in\{1,5\}} \beta_w^+ Post^w \cdot \tau_{wg}^+ + \sum_{w\in\{5,9\}} \beta_w^- Post^w \tau_{wg}^-}_{\text{dDff-in-diff estimator}} \end{aligned} \tag{B.1}$$

$$+\underbrace{\sum_{w\in\{1,5\}}\phi_w^+ R(i,w)+\sum_{w\in\{5,9\}}\phi^- R(i,w)}_{\text{Spillover Propensity}} \tag{B.2}$$

$$\begin{aligned}&+\sum_{w\{1,5\}}\mathbb{E}[\eta_{iw}^+|Post^w\cdot\tau_{wg}^+,R(i,w)]+\varepsilon_{it}.\\&\underbrace{+\sum_{w\{5,9\}}\mathbb{E}[\eta_{iw}^-|Post^w\cdot\tau_{wg}^-,R(i,w)]}_{\text{Spillover Unobserved effects}}\end{aligned} \tag{B.3}$$

The first two lines in the previous equation represent the baseline diff-in-diff model from Section 6. The third line measures the propensity to which each untreated unit is likely to be affected by spillovers arising from treated units in week w. This propensity is represented via $R(i, w)$. For robustness, we consider three different ways to capture the spillover effect propensity, $R(i, w)$:

• *Measure 1 (overlapping legs).* The first metric is R_{leg}, a binary variable indicating whether an untreated observation had an overlapping flight leg with a treated unit. The idea is that spillover effects are more likely to affect untreated ODs that share a flight leg with a treated OD. For instance, if the {Miami-Boston-London} OD was treated, then the {Miami-Boston-Paris} OD has a higher propensity to be affected by spillovers than, say, the {Miami-Atlanta-Istanbul} OD.

• *Measure 2 (overlapping airport terminals).* The second metric, $R_{airport}$, measures the fraction of airports that an untreated OD shares with all treated ODs. For example, if the {Miami-Boston-London} OD is treated, then itineraries involving the Miami, Boston, and London airports are more likely to be affected by spillovers than, say, the {Paris-Istanbul-Shanghai} OD.

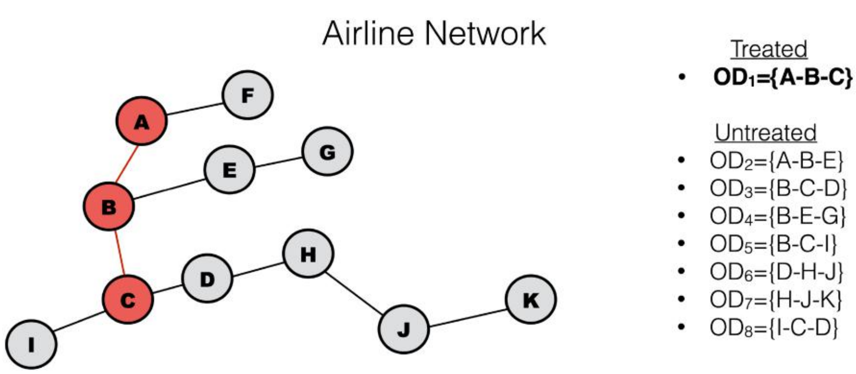
• *Measure 3 (demand correlation between ODs).* The third metric, R_{corr}, measures the degree of demand correlation between an untreated OD and all treated ODs. We measure this correlation via historical data from the year preceding the experiment. If the demand correlation between an untreated OD and a treated OD is zero (historically speaking), then the untreated OD is unlikely to be affected by the experiment. But if the historical data indicates that the demand correlation is high, then the untreated OD has a high propensity to be affected by the experiment.

To illustrate, consider Figure B.3, which shows an airfare network made up by 11 airports (A, B, C, … , K) and eight ODs traversing through these airports.

In this illustration, suppose that we treat OD_1, which flies through the route $\{A-B-C\}$, thereby affecting the fares of two flight legs (AB and BC). The figure illustrates how R_i would be constructed for each of the untreated ODs. None of these metrics is ruling out spillover effects from any of the ODs a priori, nor imposing any structure on the function $R(i, w)$. For example, the model does not assume that if $R_{leg} = 0$, then an OD is not affected by spillover effects; it also does not assume that the spillovers are increasing (or decreasing) as a function of $R(i, w)$. The functions simply establish that itineraries with a value of $R_{leg} = 1$ have a different propensity of being affected by spillovers than those with a value of $R_{leg} = 1$. To avoid imposing any structure on the nature of the spillovers, we will model this function via smoothed splines (cross-validating the regression span to maximize the functional fit).

The fourth line in the previous equation includes error terms for each of the staggered diff-in-diff shocks. These shocks are estimated for each of the support points in the propensity metric, assuming iid binomial trials. For more

Figure B.3. (Color online) Supply Chain Propensity Metrics



Spillover propensity for OD_2 -OD_8

Overlapping legs (R_{leg})

OD_2	OD_3	OD_4	OD_5	OD_6	OD_7	OD_8
1	1	0	1	0	0	0

Overlapping airports ($R_{airport}$)

OD_2	OD_3	OD_4	OD_5	OD_6	OD_7	OD_8
2/3	2/3	1/3	2/3	0	0	1/3

Demand correlation (R_{corr})

OD_2	OD_3	OD_4	OD_5	OD_6	OD_7	OD_8
$r_{1,2}$	$r_{1,3}$	$r_{1,4}$	$r_{1,5}$	$r_{1,6}$	$r_{1,7}$	$r_{1,8}$

details on the estimation of this error, see Clarke (2017) and Butts (2022).

Table B.1 shows results for the two spillover models, the leftmost column includes the estimates of the baseline model, for reference, whereas the estimates from the second, third and fourth columns show spillovers-robust estimates for metrics R_{leg}, $R_{airport}$, and R_{corr}. Because of space constraints, we only report coefficients for the staggered diff-in-diff estimates (β) and for the spillover closeness metric (ϕ).

Our estimates indicate some spillover effects across untreated ODs that have overlapping flight legs and, to some extent, across those that have correlated demand. However—as can be seen by the coefficients, ϕ—the magnitude of these spillovers is relatively small, to the point where our estimates were not visibly affected (neither in magnitude nor in statistical significance).

B.4. Variation in Departure Dates

We restricted our experimental population to itineraries that were still at the bottom of the ladder at the outset of the experiment. To this end, we focused on itineraries with far-enough departure dates, but also those that were close enough to the beginning of the "hot part" of the selling season (within 120 days of the departure date). This restriction, we argued in Section 4, allowed us to set a common starting point—that is, to consistently observe how itineraries escalate along the ladder as a function of the two pricing strategies.

Despite such common ground, there is still heterogeneity when it comes to the itineraries' departure dates. Although some itineraries had a seven-week lead time before departure, others had a nine-week lead time when the experiment began.

To determine whether this source of heterogeneity influenced our estimates, we re-estimated our results by dissecting the sample as a function of the days remaining to departure (DRD). We did so by dividing the sample into 10 buckets, according to the deciles of the DRD distribution (at the lowest decile, we find those itineraries with the fewest days remaining to departure). We then re-ran the analysis of Section 7.2 across each subsample. Figures B.4 and B.5 reproduce the exhibits from Figure 8, across each decile. There are idiosyncratic behaviors in the demand across each market, particularly at the lower end of the DRD distribution. Nevertheless, these confounding effects are not present since all the effects are benchmarked against historical demand across treatment and control pairs. Thus, idiosyncratic differences in the DRD distribution are absorbed by our model.

The plots exhibit the same pattern of the aggregate plot reported in Section 7.2. However, the pattern is slightly more pronounced across the lowest deciles. This result is expected, given that itineraries with closer departure dates were submerged in the "hot selling season" for a longer time span. The differences across each decile, however, are small in size and statistically insignificant.

Appendix C. Analytical Model of Lead-in Markups

We propose a stylized model to identify the effects of lead-in fare markups. We first consider a single-class setting to isolate the impact of a fare increase in a static setting and then extend the model to a two-class setting to identify the interplay of lead-in markups and revenue management.

C.1. Single-Class Setting

Consider two airlines, indexed by $i = 1, 2$, each with a single fare class. We denote Airline i's price by p_i, its capacity by C_i, and the quality of its product by q_i. Without loss of generality, Airline 2 has a higher-quality product than Airline 1, so $q_2 > q_1$. Let $\Delta = q_2 - q_1$ denote the quality differential.

Demand is captured by a continuum D of customers. Let $\mathcal{U}_i$ denote the utility of a customer who purchases a ticket from Airline i. We consider the following linear utility function:

$$\mathcal{U}_i = \overline{U} - p_i + \theta q_i,$$

where $\overline{U}$ is a base utility level and θ captures customers' quality sensitivity. We assume that $\overline{U}$ is large enough so that all customers choose to purchase a ticket from one airline—that is, there is no outside option. We also assume that total capacity is sufficient to accommodate customer demand, i.e., $C_1 + C_2 \geq D$. Customers are utility maximizers, so each customer prefers to purchase a ticket from Airline 2 if $\theta \geq \frac{p_2 - p_1}{\Delta}$, and from Airline 1 otherwise.[5]

To capture customer heterogeneity, we represent θ by means of a probability distribution F. A typical segmentation in the airline industry distinguishes leisure travelers—who often make booking decisions based on prices—and business travelers—who are often willing to pay a premium for higher-quality itineraries. Accordingly, we assume that F has an atom $1-\mu$ in $\theta = 0$, that is, a mass $1-\mu$ of customers choose the lower-price option. Conditionally on $\theta > 0$, θ is uniformly distributed between 0 and $\overline{\theta}$. The parameters μ and $\overline{\theta}$ capture customers' quality sensitivity—μ measures the proportion of quality-sensitive customers, and $\overline{\theta}$ measures their willingness to pay.

We now compare the effects of lead-in matching versus markup for Airline 2. Under lead-in matching, we have $p_2 = p_1$. Under a lead-in markup, we write $p_2 = \beta p_1$, with $\beta > 1$. We assume that $\frac{p_2 - p_1}{\Delta} \leq \overline{\theta}$ (otherwise, Airline 2's product is clearly overpriced).

- Under matching, Airline 2 sells $\min(D, C_2)$ tickets at price p_1. Let π_2^M be its revenue. We have

$$\pi_2^M = p_1 \times \min(D, C_2). \tag{C.1}$$

- Under markup, Airline 2 would capture $D\left(1 - F\left(\frac{p_2 - p_1}{\Delta}\right)\right) = D\mu\left(1 - \frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right)$ customers without capacity restrictions. If this quantity exceeds C_2, Airline 2 only sells C_2 tickets. Vice versa, if residual demand exceeds Airline 1's capacity, Airline 1 only sells C_1 tickets and Airline 2 sells $D - C_1$ tickets. Airline 2's revenue, denoted by π_2^D (D for "differentiation"), is thus

$$\pi_2^D = \beta p_1 \times \max\left\{\min\left(D\mu\left(1 - \frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right), C_2\right), D - C_1\right\}. \tag{C.2}$$

In the single-class setting, a lead-in markup (weakly) reduces Airline 2's sales and increases its yield—Airline 2 sells fewer tickets at a higher price.

Remark C.1. Airline 2's sales are (weakly) lower under a lead-in markup than under lead-in matching. If $C_2 \leq D\mu\left(1 - \frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right)$, Airline 2's sells C_2 seats across the two strategies. Otherwise, Airline 2's sales are strictly lower under a lead-in markup than under lead-in matching.

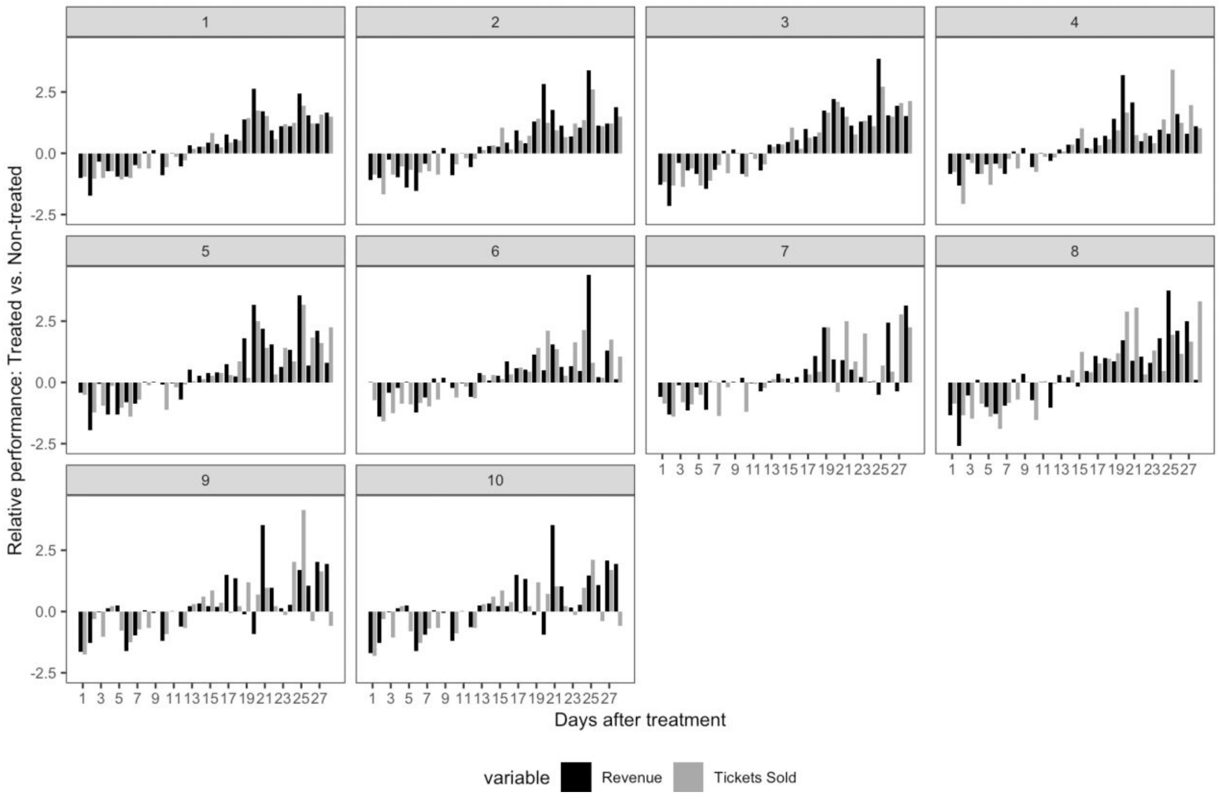
Table B.1. Network Spillover Effects

	Revenue				Yield				Tickets sold				Market share			
Dependent variable	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
β_1^+	0.21***	0.19**	0.21***	0.21***	0.22***	0.20***	0.22***	0.22***	−0.10	−0.10	−0.10	−0.11	−0.13	−0.13	−0.13	−0.13
	(0.06)	(0.09)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.15)	(0.15)	(0.15)	(0.15)	(0.13)	(0.13)	(0.13)	(0.13)
β_5^-	−0.20***	−0.18*	−0.20***	−0.20***	−0.21**	−0.18*	−0.21**	−0.21**	0.18	0.18	0.18	0.12	−0.10	−0.10	−0.10	−0.10
	(0.08)	(0.10)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.19)	(0.19)	(0.19)	(0.19)	(0.16)	(0.16)	(0.16)	(0.16)
β_5^+	0.23***	0.21***	0.23***	0.21***	0.23***	0.18***	0.22***	0.21***	-0.58*	-0.58*	-0.58*	-0.58*	0.02	0.02	0.02	0.02
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.35)	(0.35)	(0.36)	(0.38)	(0.12)	(0.12)	(0.12)	(0.12)
β_9^-	−0.24**	−0.21**	−0.23**	−0.23**	−0.24*	−0.21*	−0.24*	−0.24*	0.14	0.14	0.14	0.12	0.00	0.00	0.00	0.00
	(0.12)	(0.10)	(0.13)	(0.13)			(0.14)	(0.14)	(0.31)	(0.31)	(0.31)	(0.31)	(0.26)	(0.26)	(0.26)	(0.26)
ϕ_1^+		0.01*	0.00	0.03		0.00	0.00	0.00		−0.02*	0.00	−0.03*		0.00	0.00	0.01
		(0.006)	(0.01)	(0.03)		(0.00)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)		(0.03)	(0.01)	(0.01)
ϕ_5^-		−0.02**	0.00	0.01		0.02*	0.00	0.00		0.00	0.00	0.00		0.00	0.00	0.00
		(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)		(0.04)	(0.01)	(0.01)
ϕ_5^+		0.01	0.00	0.00		0.00	0.01	0.01		0.00	0.00	0.02		0.00	0.00	0.00
		(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)		(0.02)	(0.01)	(0.01)		(0.05)	(0.01)	(0.01)
ϕ_9^-		0.01	0.01	0.02		-0.01	-0.01	0.00		0.01*	0.00	0.02		0.00	0.01	0.01
		(0.007)	(0.03)	(0.03)		(0.01)	(0.01)	(0.01)		(0.001)	(0.00)	(0.00)		(0.10)	(0.15)	(0.15)
Itineraries	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280	13,280
Closeness metric	None	R_{leg}	$R_{terminal}$	R_{corr}	None	R_{leg}	$R_{terminal}$	R_{corr}	None	R_{leg}	$R_{terminal}$	R_{corr}	None	R_{leg}	$R_{terminal}$	R_{corr}

Notes. The specifications in Columns I, V, IX, and XIII correspond to the baseline model. The specifications in Columns II, VI, X, and XIV correspond to spillovers-robust estimates with the R_{leg} metric. The specifications in Columns III, VII, XI, and XV correspond to spillover-robust estimates with the $R_{airport}$ metric. The specifications in Columns IV, VIII, XII, and XVI correspond to spillovers-robust estimates with the R_{corr} metric. The table reports the estimated coefficient values and standard errors (in parentheses) of the diff-in-diff estimates in Equation (2).

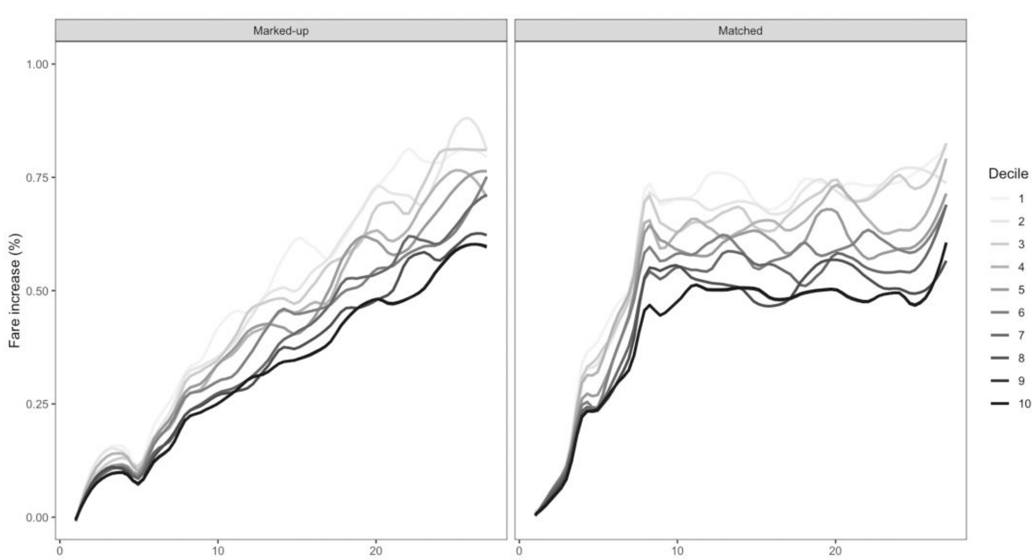
*, **, and ***Significance levels at 10%, 5%, and 1%, respectively.

Figure B.4. Revenue and Sales Dynamics, Dissected per Decile of the "Days Remaining to Departure" Distribution



Notes. This exhibit reproduces Figure 8 (left), broken down into each decile of the DRD distribution. The *y* axis displays the revenue and sales of treated itineraries, relative to control group itineraries, in terms of standard deviations. If the bar is positive, treated itineraries outperformed the control group itineraries. The *x* axis reports the days since the beginning of the treatment.

Figure B.5. Percentage Fare Increase During the Treated Days per Decile of the "Days Remaining to Departure" Distribution



Notes. This exhibit reproduces Figure 8 (right) but divided into two plots (each containing a given pricing strategy). The *y* axis displays the fare percentage increase of marked-up and control itineraries, broken down into each decile of the DRD distribution. The *x* axis reports the days since the beginning of the treatment.

Remark C.2. A lead-in markup increases Airline 2's yield by a factor β.

Proposition C.1 shows that the net effect of a lead-in markup on Airline 2's revenue can be positive.

Proposition C.1. *In the single-class setting, a lead-in markup increases Airline 2's revenue if*

$$\beta\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right)\geq 1 \tag{C.3}$$

The proposition provides a sufficient condition under which a unilateral lead-in markup increases Airline 2's revenue. This condition reveals that a lead-in markup is more beneficial as the quality differential becomes stronger and as customers become more quality-sensitive (Equation (C.3) is more likely to be satisfied as Δ, μ and $\overline{\theta}$ get larger). Moreover, a lead-in markup is more beneficial with an intermediate price differential. To see this, note that Equation (C.3) is quadratic in β, which implies that there exist $\underline{\beta}$ and $\overline{\beta}$ such Airline 2's revenue increases as long as $\underline{\beta}\leq\beta\leq\overline{\beta}$. The intuition behind this result is that, if β is too small, the price differential is not sufficient to offset the loss of price-sensitive customers. If β is too large, Airline 2 loses too many customers, which is not compensated by the price increase. In-between, Airline 2 may benefit from a moderate differential.

C.2. Two-Class Setting

To reflect revenue management dynamics, we now assume that each airline has two fare classes: Class A (lead-in) and Class B (sell-up). We index them by $j=A,B$. Class A customers arrive before Class B customers Let L_i be Airline i's booking limit (maximum number of Class A tickets). We assume that the booking limits are exogenous.

We denote Airline i's fare in Class j by p_{ji}. Lead-in matching means that $p_{A2}=p_{A1}$ and $p_{B2}=p_{B1}$. Under lead-in markups, we assume that the relative price differential is identical across the two classes but this does not drive our results. We write $p_{A2}=\beta p_{A1}$ and $p_{B2}=\beta p_{B1}$, with $\beta>1$.

A mass D_j of customers is looking to buy Class j tickets, for $j=A,B$. As in the baseline setting, total capacity is sufficient to accommodate customer demand, i.e., $C_1+C_2\geq D$. We also assume that all Class A demand can be accommodated across the two airlines, that is, $L_1+L_2\geq D_A$.

Let $\mathcal{U}_{ji}$ denote the utility of a Class j customer purchasing a ticket from Airline i. It is given by

$$\mathcal{U}_{ji}=\overline{U}_j-p_{ji}+\theta_j q_i.$$

This setting relates to the model from Netessine and Shumsky (2005), who study a revenue management game between two competing airlines—by optimizing each airline's booking limit. The authors also consider sequential arrivals of Class A and Class B customers, but assume that each airline faces its own demand (each customer attempts to purchase a ticket from one airline and, if rejected, turns to the other airline). In contrast, our model considers an aggregate demand, whose distribution across the two airlines is endogenous to their price and quality.

As in the baseline model, we assume that $\overline{U}_A$ and $\overline{U}_B$ are large enough so that all customers purchase a ticket from one airline. Moreover, we assume that $\overline{U}_B-\overline{U}_A$ is large enough so that there is no sell-up—that is, customers looking to purchase a Class A ticket will not purchase a Class B ticket. Each Class j customer prefers Airline 2's product if and only if $\theta_j\geq\frac{p_{j2}-p_{j1}}{\Delta}$.

The parameter θ_j still follows a probability distribution. A mass $1-\mu_j$ of Class j customers are price-sensitive. For the remaining mass μ_j, θ_j is uniformly distributed between 0 and $\overline{\theta}_j$. Under lead-in matching, a mass D_j of Class j customers prefers Airline 2's product over Airline 1's. Under lead-in markups, a mass $D_j\mu_j\left(1-\frac{(\beta-1)p_{j1}}{\overline{\theta}_j\Delta}\right)$ of Class j customers prefers Airline 2's product. However, Airline 2's sales depend on the booking limits L_1, L_2, and the capacities C_1, C_2.

We denote by π_2^M and π_2^D the revenue of Airline 2 under the lead-in matching strategy and lead-in markup, respectively. By proceeding as in Equations (C.1) and (C.2), we get

$$\pi_2^M=p_{A1}\times\min(D_A,L_2)+p_{B1}\times\min(D_B,C_2-\min(D_A,L_2)), \tag{C.4}$$

$$\pi_2^D=\beta p_{A1}\times\max\left\{\min\left(D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right),L_2\right),D_A-L_1\right\}$$
$$+\beta p_{B1}\times\max\left\{\min\left(D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right),\overline{C}_{B2}\right),D_B-\overline{C}_{B1}\right\}, \tag{C.5}$$

where $\overline{C}_{Bi}$ denotes the remaining capacity of Airline i after Class A sales. In particular,

$$\overline{C}_{B2}=C_2-\max\left\{\min\left(D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right),L_2\right),D_A-L_1\right\}.$$

Proposition C.2 identifies a sufficient condition under which a lead-in markup increases Airline 2's revenue—extending Proposition C.1 to each class. As in the single-fare setting, moderate lead-in markups can be beneficial, especially under strong quality differentiation and quality-sensitive customers. Importantly, the condition does not depend on the airlines' capacities and booking limits—lead-in markups increase Airline 2's revenue regardless of booking limits and capacities.

Proposition C.2. *In the two-class setting, a lead-in markup increases Airline 2's revenue if:*

$$\beta\mu_j\left(1-\frac{(\beta-1)p_{j1}}{\overline{\theta}_j\Delta}\right)\geq 1,\quad \textit{for } j=A,B. \tag{C.6}$$

Next, lead-in markups still have a (weakly) negative impact on Airline 2's sales. Indeed, fewer customers seek to purchase a ticket from Airline 2 when it charges higher prices—so Airline 2's sales decrease when capacity constraints are inactive or remain unchanged otherwise.

Remark C.3. Airline 2's sales are (weakly) lower under lead-in markups than under matching.

Last, lead-in markups do not necessarily increase Airline 2's yield. In other words, higher fares may result in a lower average price per ticket—because markups can change the customer mix.

Remark C.4. Airline 2's yield is generally larger under lead-in markups than under lead-in matching; however, there exist cases where lead-in markups result in a lower yield for Airline 2.

C.3. Discussion

The stylized model predicts that unilateral markups can increase an airline's revenue—especially if the quality differential is large and if customers are sufficiently quality-sensitive. In these cases, the lead-in markup needs to be large enough to offset the loss of price-sensitive customers, yet small enough to attract a sufficient mass of quality-sensitive customers.

Recall that Equation (C.3) provides a sufficient condition for the lead-in markup to be beneficial for Airline 2, by increasing its revenue in each fare class. But Airline 2 can also benefit from a lead-in markup when this condition is not satisfied; for instance, a lead-in markup can decrease the revenue from Class A but increase the total revenue. Unfortunately, the analysis is quite intricate and depends on the various parameters in the model. To avoid enumerating all possible cases without deep insights, let us focus on one typical regime for simplicity.

Specifically, assume that, under a matching strategy, Airline 2 sells Class A tickets (at the lead-in fare) up to the booking limit and Class B tickets (at the sell-up fare) up to the aircraft's capacity. This is a reasonable regime since Airline 2 has the higher-quality itinerary, so we can expect the capacity constraints to be binding. Under a markup strategy, Airline 2 will face a lower effective demand. To be conservative, let us assume that its booking limit and the capacity of its aircraft will not be binding—otherwise, Airline 2's revenue is clearly higher under a markup strategy than under a matching strategy. Therefore, under a matching strategy, Airline 2 sells L_2 tickets in Class A and $C_2 - L_2$ tickets in Class B. Under a markup strategy, Airline 2 sells $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right)$ tickets in Class A and $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right)$ tickets in Class B.

In that case, demand substitution occurs if the following holds:

$$\text{Class A} \qquad D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) < L_2$$

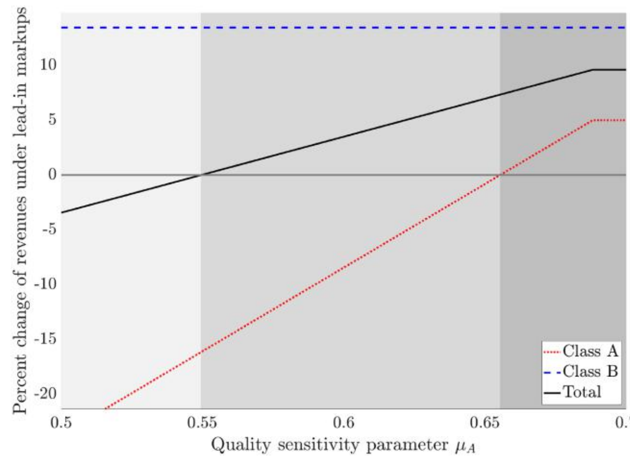
$$\text{Class B} \qquad D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) > C_2 - L_2.$$

The following condition is sufficient to ensure an increase in the airline's overall revenue:

$$\frac{p_{B1}}{p_{A1}}\left\{\beta D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) + L_2 - C_2\right\} \geq L_2 - \beta D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right).$$

This condition confirms the qualitative findings highlighted previously—Namely, lead-in markups are more beneficial when the quality differential is strong, when the price differential is moderate, and when customers are more quality sensitive. However, it also gives rise to four new factors driving a revenue increase: the sell-up rules (i.e., p_{B1} versus p_{A1}), the aircraft's capacities (i.e., C_1 and C_2), the revenue management rules (i.e., L_1 and L_2), and the demand dynamics (i.e., D_A and D_B).

Figure C.1. (Color online) Airline 2's Revenue as a Function of Quality Sensitivity of Class A Customers



Note. Parameters: $D_A = 80, D_B = 85, \Delta = 100$, $p_{A1} = \$1,000$, $p_{A2} = \$1,200$, $\beta = 1.05, L_1 = L_2 = 55, C_1 = C_2 = 130, \mu_B = 0.7, \overline{\theta}_A = \overline{\theta}_B = 500$.

To illustrate this situation, Figure C.1 plots the impact of a lead-in markup on Airline 2's revenue, as a function of the proportion μ_A of quality-sensitive Class A customers. We assume that $\mu_A \leq \mu_B$: Class B customers are more quality-sensitive than Class A customers. The figure highlights three regimes. When μ_A is large, a lead-in markup increases Airline 2's revenue in each fare class. When μ_A is small, a lead-in markup induces large losses in Class A revenue, resulting in a decrease in Airline 2's total revenue. In between, a lead-in markup reduces Airline 2's Class A revenue but an increase its total revenue—shifting the distribution of sales toward the higher class.

To derive these insights, the model has made a number of assumptions—for instance, by considering two customer segments, a simple model of customer choice, and two fare classes. Moreover, the model involves parameters that are very hard to calibrate and estimate in practice—such as the quality differential (Δ) and customers' quality sensitivity (μ and θ). It is thus difficult to assess, from this model, the ultimate impact of lead-in markups on an airline's revenue, yield, and sales.

These limitations motivate our field experiment to assess these effects in practice. Our experimental results show that lead-in markups can increase revenue and yield without negatively impacting sales and market share. Moreover, our empirical results confirm the mechanism of demand substitution induced by lead-in markups, by shifting the distribution of sales toward the end of the selling season, and identify boundary conditions on the benefits of lead-in markups by characterizing the drivers of quality sensitivity in the airline industry.

C.4. Proof of Statements

Proof of Proposition C.1. We distinguish two cases.

Case 1: $D \leq C_2$. Under lead-in matching, Airline 2 sells D tickets and $\pi_2^M = p_1 D$. Because $D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right) \leq D \leq C_2$, Airline 2's capacity is not binding under a lead-in markup either, so $\pi_2^D = \beta p_1 \max\left\{D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right), D - C_1\right\} \geq \beta p_1 D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right)$. From Equation (C.3), we get $\pi_2^D \geq \pi_2^M$.

Case 2: $D > C_2$. Under lead-in matching, Airline 2 sells C_2 tickets and $\pi_2^M = p_1 C_2$.

- If $D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right) \geq C_2$, then $\pi_2^D = \beta p_1 C_2 > p_1 C_2 = \pi_2^M$.
- Otherwise, $\pi_2^D = \beta p_1 \max\left\{D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right), D - C_1\right\} \geq \beta p_1 D\mu\left(1-\frac{(\beta-1)p_1}{\overline{\theta}\Delta}\right)$. From Equation (C.3), we obtain $\pi_2^D \geq p_1 D > p_1 C_2 = \pi_2^M$.

Proof of Proposition C.2. We distinguish four cases.

- *Case 1:* $D_A \leq L_2$ and $D_B \leq C_2 - D_A$.

In this case, we have $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \leq L_2$ and $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \leq C_2 - D_A \leq \overline{C}_{B2}$. It becomes

$$\pi_2^M = p_{A1}D_A + p_{B1}D_B,$$

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} \\ &\quad + \beta p_{B1}\max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &\geq \beta p_{A1}D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) + \beta p_{B1}D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \\ &\geq p_{A1}D_A + p_{B1}D_B \ \text{ from Equation (C.6).}\end{aligned}$$

- *Case 2:* $D_A \leq L_2$ *and* $D_B > C_2 - D_A$.

We have

$$\pi_2^M = p_{A1}D_A + p_{B1}(C - D_A) < p_{A1}D_A + p_{B1}D_B.$$

We know that $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \leq D_A \leq L_2$, so that Airline 2 can satisfy all demand from Class A customers under lead-in markups (because $D_A - L_1 \leq L_2$). We distinguish two subcases:

- *Case 2.1:* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \leq \overline{C}_{B2}$.

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} \\ &\quad + \beta p_{B1}\max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &\geq \beta p_{A1}D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) + \beta p_{B1}D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \\ &\geq p_{A1}D_A + p_{B1}D_B \quad \text{from Equation (C.6)} \\ &> \pi_2^M.\end{aligned}$$

- *Case 2.2:* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) > \overline{C}_{B2}$.

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times\max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} \\ &\quad + \beta p_{B1}\times\left[C_2 - \max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\}\right].\end{aligned}$$

If $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \leq D_A - L_1$, we obtain

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times D_A + \beta p_{B1}\times(C - D_A) \\ &\quad + \left((1-\mu_A) + \beta(p_{B1} - p_{A1})\mu_A\frac{\beta p_{A1}}{\Delta}\right)D_A \\ &= \beta\pi_2^M + \left((1-\mu_A) + \mu_A\frac{\beta p_{A1}}{\Delta}\right)D_A\beta(p_{B1} - p_{A1}) \\ &> \pi_2^M.\end{aligned}$$

Otherwise,

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times D_A + \beta p_{B1}\times(C - D_A) + \beta(p_{B1} - p_{A1})L_1 \\ &= \beta\pi_2^M + \beta(p_{B1} - p_{A1})L_1 \\ &> \pi_2^M.\end{aligned}$$

- *Case 3:* $D_A > L_2$ and $D_B \leq C_2 - L_2$.

We have

$$\pi_2^M = p_{A1}L_2 + p_{B1}D_B < p_{A1}D_A + p_{B1}D_B.$$

We know that $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \leq D_B \leq C_2 - L_2$, so that Airline 2 can satisfy all demand from Class B customers under lead-in markups. We distinguish two subcases:

- *Case 3.1:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \geq L_2$. We have

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times L_2 + \beta p_{B1}\times\max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &\geq \beta p_{A1}\times L_2 + \beta p_{B1}\times D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \\ &> p_{A1}L_2 + p_{B1}D_B \ \text{ from Equation (C.6)} \\ &= \pi_2^M.\end{aligned}$$

- *Case 3.2:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) < L_2$. We have

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times\max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} \\ &\quad + \beta p_{B1}\times\max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &\geq \beta p_{A1}\times D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \\ &\quad + \beta p_{B1}\times D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \\ &\geq p_{A1}D_A + p_{B1}D_B \quad \text{from Equation (C.6)} \\ &> \pi_2^M.\end{aligned}$$

- *Case 4:* $D_A > L_2$ and $D_B > C_2 - L_2$. We have

$$\pi_2^M = p_{A1}L_2 + p_{B1}(C_2 - L_2) < p_{A1}D_A + p_{B1}D_B.$$

We distinguish four subcases:

- *Case 4.1:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \leq L_2$ *and* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \leq \overline{C}_{B2}$.

$$\begin{aligned}\pi_2^D &= \beta p_{A1}\times\max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} \\ &\quad + \beta p_{B1}\times\max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &\geq \beta p_{A1}\times D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) + \beta p_{B1} \\ &\quad \times D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \\ &\geq p_{A1}D_A + p_{B1}D_B \quad \text{from Equation (C.6)} \\ &> \pi_2^M.\end{aligned}$$

- *Case 4.2:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \le L_2$ *and* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) > \overline{C}_{B2}$.

$$\pi_2^D = \beta p_{A1} \times \max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\} + \beta p_{B1} \times \left[C_2 - \max\left\{D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right), D_A - L_1\right\}\right].$$

If $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \le D_A - L_1$, we obtain

$$\begin{aligned}\pi_2^D &= \beta p_{A1} \times L_2 + \beta p_{B1} \times (C_2 - L_2) + \beta(p_{B1} - p_{A1}) \\ &\quad \times \left[L_2 - D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right)\right] \\ &= \beta\pi_2^M + \beta(p_{B1} - p_{A1}) \times \left[L_2 - D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right)\right] \\ &> \pi_2^M.\end{aligned}$$

Otherwise,

$$\begin{aligned}\pi_2^D &= \beta p_{A1} \times L_2 + \beta p_{B1} \times (C - L_2) + \beta(p_{B1} - p_{A1})(L_1 + L_2 - D_A) \\ &= \beta\pi_2^M + \beta(p_{B1} - p_{A1})(L_1 + L_2 - D_A) \\ &> \pi_2^M.\end{aligned}$$

- *Case 4.3:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) > L_2$ *and* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \le \overline{C}_{B2}$.

$$\begin{aligned}\pi_2^D &= \beta p_{A1} \times L_2 + \beta p_{B1} \times \max\left\{D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right), D_B - \overline{C}_{B1}\right\} \\ &> p_{A1}L_2 + p_{B1}D_B \quad \text{from Equation (C.6)} \\ &> \pi_2^M.\end{aligned}$$

- *Case 4.4:* $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) > L_2$ *and* $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) > \overline{C}_{B2}$.

$$\pi_2^D = \beta p_{A1} \times L_2 + \beta p_{B1} \times (C_2 - L_2) = \beta\pi_2^M > \pi_2^M.$$

Proof of Remark C.4. We seek an example where Airline 2's yield is higher under lead-in matching than under lead-in markups. Consider Case 3.1 from the proof of Proposition C.2, that is, $D_A > L_2$, $D_B \le C_2 - L_2$, and $D_A\mu_A\left(1-\frac{(\beta-1)p_{A1}}{\overline{\theta}_A\Delta}\right) \ge L_2$. Assume further that $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right) \le D_B - \overline{C}_{B1}$. In this case, Airline 2 sells L_2 Class A tickets under lead-in matching and under lead-in markups. Airline 2 sells D_B Class B tickets under lead-in matching and $D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right)$ Class B tickets under lead-in markups. Lead-in markups can decrease Airline 2's yield if

$$\frac{p_{A1}L_2 + p_{B1}D_B}{L_2 + D_B} > \beta \times \frac{p_{A1}L_2 + p_{B1}D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right)}{L_2 + D_B\mu_B\left(1-\frac{(\beta-1)p_{B1}}{\overline{\theta}_B\Delta}\right)}.$$

Endnotes

[1] The same result could be achieved by modifying the revenue management algorithms, that is, by introducing quality considerations into the revenue management algorithms (i.e., the bid-price or the quantity controls). Although we do not formally explore this mechanism in our paper, we discuss this alternative strategy in Section 8.

[2] To clarify, an inventory class is different from a seat class, in that (i) the seat class refers to the cabin service (e.g., first, business, and economy), whereas (ii) the inventory class is a fare hierarchy. For example, inventory classes $(A, \ldots, G)$ could belong to the economy seat class whereas inventory classes (H, I) to the business class. Also, the premium seat classes (business and first class) are priced independently, meaning that they are not anchored to the lead-in fare. Finally, airlines sometimes create subclasses within the same cabin—for example, "basic economy" or "premium economy." When it comes to pricing the ladder, however, the lead-in fare is the lowest-possible fare offered in the economy cabin (regardless of these subclasses).

[3] The fare changes on the rightmost plot exhibit an upward trend but are not strictly monotonic. The non-monotonicity is because of variations in the population of itineraries purchased at different times of the selling season. Although the price path of each itinerary is generally increasing over time, the average price may not be monotonically increasing if more tickets are purchased on expensive itineraries on a given day and fewer tickets are purchased on cheaper itineraries the following day.

[4] The causal forest approach has a potential pitfall, in that the estimation is susceptible to unobserved confounders. Wager and Athey (2018) discuss the influence of these confounders in the spirit of a natural-experiment setting, where matching is done in the absence of randomization (or quasi-randomization). This issue, however, is not prominent in our paper because the treatment and control groups were randomized within the experimental population.

[5] If $p_1 = p_2$, all customers prefer to purchase Airline 2's product due to quality differentiation, even if $\theta = 0$.

References

Adida E, Perakis G (2010) Dynamic pricing and inventory control: Uncertainty and competition. *Oper. Res.* 58(2):289–302.

Angelucci M, Di Maro V (2016) Programme evaluation and spillover effects. *J. Development Effect* 8(1):22–43.

Angrist JD, Pischke J-S (2008) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Austin PC (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.* 46(3):399–424.

Austin PC, Stuart EA (2015) Moving toward best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statist. Medicine* 34(28):3661–3679.

Belobaba P, Odoni A, Barnhart C (2015) *The Global Airline Industry* (John Wiley & Sons, Hoboken, NJ).

Belobaba PP (1989) Application of a probabilistic decision model to airline seat inventory control. *Oper. Res.* 37(2):183–197.

Belobaba PP, Wilson JL (1997) Impacts of yield management in competitive airline markets. *J. Air Transportation Management* 3(1):3–9.

Berg T, Reisinger M, Streitz D (2020) Handling spillover effects in empirical research. *J. Financial Econom.* 142(3):1109–1127.

Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* 5(3):203–229.

Bray RL, Coviello D, Ichino A, Persico N (2016) Multitasking, multiarmed bandits, and the Italian judiciary. *Manufacturing Service Oper. Management* 18(4):545–558.

Butts K (2022) Difference-in-differences with spatial spillovers. Working paper.

Cameron AC, Miller DL (2015) A practitioner's guide to cluster-robust inference. *J. Human Resources* 50(2):317–372.

Caro F, Gallien J (2012) Clearance pricing optimization for a fast-fashion retailer. *Oper. Res.* 60(6):1404–1422.

Clarke D (2017) Estimating Difference-in-differences in the presence of spillovers. Working paper, University of Santiago, Chile.

Clarke D (2019) Estimating difference-in-differences in the presence of spillovers. MPRA Paper 81604, University Library of Munich, Germany.

Feng Y, Gallego G (1995) Optimal starting times for end-of-season sales and optimal stopping times for promotional fares. *Management Sci.* 41(8):1371–1391.

Fisher M, Gallino S, Li J (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Sci.* 64(6):2496–2514.

Gallego G, Hu M (2014) Dynamic pricing of perishable assets under competition. *Management Sci.* 60(5):1241–1259.

Gallego G, Van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8):999–1020.

Gaur V, Fisher ML (2005) In-store experiments to determine the impact of price on sales. *Production Oper. Management* 14(4):377–387.

Goodman-Bacon A (2018) Public insurance and mortality: Evidence from Medicaid implementation. *J. Political Econom.* 126(1):216–262.

Koppelman FS, Coldren GM, Parker RA (2008) Schedule delay impacts on air-travel itinerary demand. *Transportation Res. Part B: Methodological* 42(3):263–273.

Lechner M (2011) The estimation of causal effects by difference-in-difference methods. *Foundations Trends® Econometrics* 4(3):165–224.

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

Lurkin V, Garrow LA, Higgins MJ, Newman JP, Schyns M (2017) Accounting for price endogeneity in airline itinerary choice models: An application to continental US markets. *Transportation Res. Part A Policy Practice* 100:228–246.

Martínez-de-Albéniz V, Talluri K (2011) Dynamic price competition with fixed capacities. *Management Sci.* 57(6):1078–1093.

McGill J, Van Ryzin G (1999) Revenue management: Research overview and prospects. *Transportation Sci.* 33(2):233–256.

Monfared S., Pavlov A. (2019) Political risk affects real estate markets. *J. Real Estate Finance and Econom.* 58:1–20.

Moreno A, Terwiesch C (2015) Pricing and production flexibility: An empirical analysis of the US automotive industry. *Manufacturing Service Oper. Management* 17(4):428–444.

Netessine S, Shumsky RA (2005) Revenue management games: Horizontal and vertical competition. *Management Sci.* 51(5): 813–831.

Nicolae M, Arikan M, Deshpande V, Ferguson M (2016) Do bags fly free? An empirical analysis of the operational implications of airline baggage fees. *Management Sci.* 63(10):3187–3206.

Olivares M, Cachon GP (2009) Competing retailers and inventory: An empirical investigation of General Motors' dealerships in isolated US markets. *Management Sci.* 55(9):1586–1604.

Perakis G, Sood A (2006) Competitive multi-period pricing for perishable products: A robust optimization approach. *Math. Programming* 107(1–2):295–335.

Ramdas K, Williams J, Lipson M (2013) Can financial markets inform operational improvement efforts? Evidence from the airline industry. *Manufacturing Service Oper. Management* 15(3): 405–422.

Rubin DB (2001) Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services Outcomes Res. Methodology* 2(3–4):169–188.

Stamatopoulos I, Bassamboo A, Moreno A (2017) The effects of menu costs on supply chain efficiency: Evidence from adoption of the electronic shelf label technology. Working paper, The University of Texas at Austin, Austin, TX.

Stuart EA, Lee BK, Leacy FP (2013) Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *J. Clinical Epidemiology* 66(8):S84–S90.

Talluri K, Van Ryzin G (1998) An analysis of bid-price controls for network revenue management. *Management Sci.* 44(11):1577–1593.

Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* 50(1):15–33.

Talluri KT, Van Ryzin GJ (2006) *The Theory and Practice of Revenue Management*, vol. 68 (Springer Science & Business Media, Boston).

Talluri KT, Van Ryzin G, Van Ryzin G (2004) *The Theory and Practice of Revenue Management*, vol. 1 (Springer, Berlin).

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113(523):1228–1242.

Welch A (2012) Quality of service index. Accessed May 6, 2022, http://www.aci-na.org/sites/default/files/welch_qsi_fundamentals.pdf.

Zhang D, Adelman D (2009) An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Sci.* 43(3):381–394.

Zhang D, Weatherford L (2017) Dynamic pricing for network revenue management: A new approach and application in the hotel industry. *INFORMS J. Comput.* 29(1):18–35.