Paying More for a Shorter Flight? - Hidden City Ticketing

Quanquan Liu

February 2020

Abstract

Hidden city ticketing occurs when an indirect flight from city A to city C through connection node city B is cheaper than the direct flight from city A to city B. Then passengers traveling from A to B have an incentive to purchase the ticket from A to C but get off the plane at B. In this paper, I build a structural model to explain the cause and impact of hidden city ticketing. I collect empirical data from the Skiplagged webpage and apply global optimization algorithms to estimate the parameters of my model. I also conduct counterfactual analysis to shed some light on policy implications. I find that hidden city opportunity occurs only when airlines are applying a hub-and-spoke network structure, under which they intend to lower their flying costs compared to a fully connected network. I find that in the short run, hidden city ticketing does not necessarily decrease airlines' expected profits. Consumer welfare and total surplus always increase. In the long run, the welfare outcomes become more complicated. For some routes airlines have the incentive to switch from hub-and-spoke network to a fully connected one when there are more and more passengers informed of hidden city ticketing. During this process, firms always result in lower expected revenue, while consumers and the whole society are not necessarily better off.

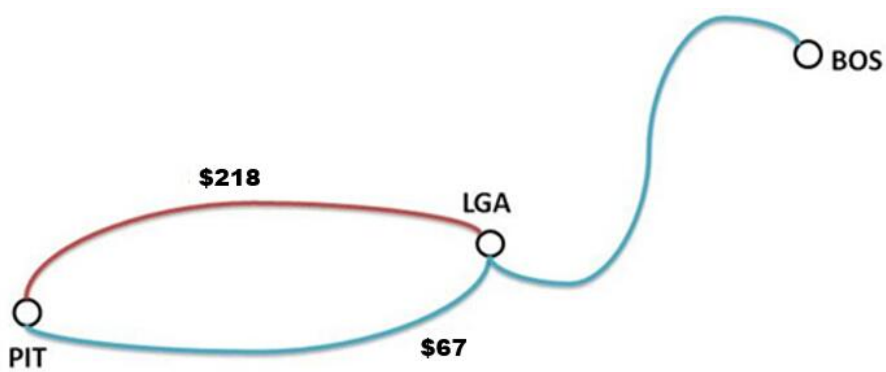
Keywords: hidden city ticketing, network structure, second-degree price discrimination, informed consumers.

1 Introduction

Hidden city ticketing is an interesting pricing phenomenon occurring after the deregulation of the airline industry in 1978 (Wang and Ye 2015). It is an airline booking strategy passengers use to reduce their flying costs. Hidden city ticketing occurs when an indirect flight from city A to city C, using city B as the connection node, turns out to be cheaper than the direct flight from city A to city B. In which case passengers who wish to fly from A to B have an incentive to purchase the indirect flight ticket, pretend to fly to city C, while disembark at the connection node B, and discard the remaining segment B to C. When this happens, city B is called the "hidden city", and this behavior is then called "hidden city ticketing".

The following real world example (Figure 1) illustrates hidden city ticketing. On November 19, 2018, a direct flight operated by Delta Air Lines flying from Pittsburgh to New York city cost $218. On the same day, for the same departure and landing time, another indirect flight also operated by Delta Air Lines flying from Pittsburgh to Boston, with one stop at New York city, cost only $67. These two flights share exactly

Figure 1: An example of hidden city ticketing.



the same first segment, while the price of the indirect one accounted for only 30% of that of the direct one. That is, you are able to fly more than 200 miles further but pay $151 less! The New York city is then called a "hidden city" in this case.

Although technically legal, hidden city ticketing actually violates the airfare rules of most airline companies in United States. For example, according to the Contract of Carriage of United Airlines (revised by December 31, 2015):

"Fares apply for travel only between the points for which they are published. Tickets may not be purchased and used at fare(s) from an initial departure point on the Ticket which is before the Passenger's actual point of origin of travel, or to a more distant point(s) than the Passenger's actual destination being traveled even when the purchase and use of such Tickets would produce a lower fare. This practice is known as "Hidden Cities Ticketing" or "Point Beyond Ticketing" and is prohibited by UA."

Passengers might be penalized when conducting hidden city ticketing. Airlines are able to "confiscate any unused Flight Coupons", "delete miles in the passenger's frequent flyer account", "assess the passenger for the actual value of the ticket", or even "take legal action with respect to the passenger".

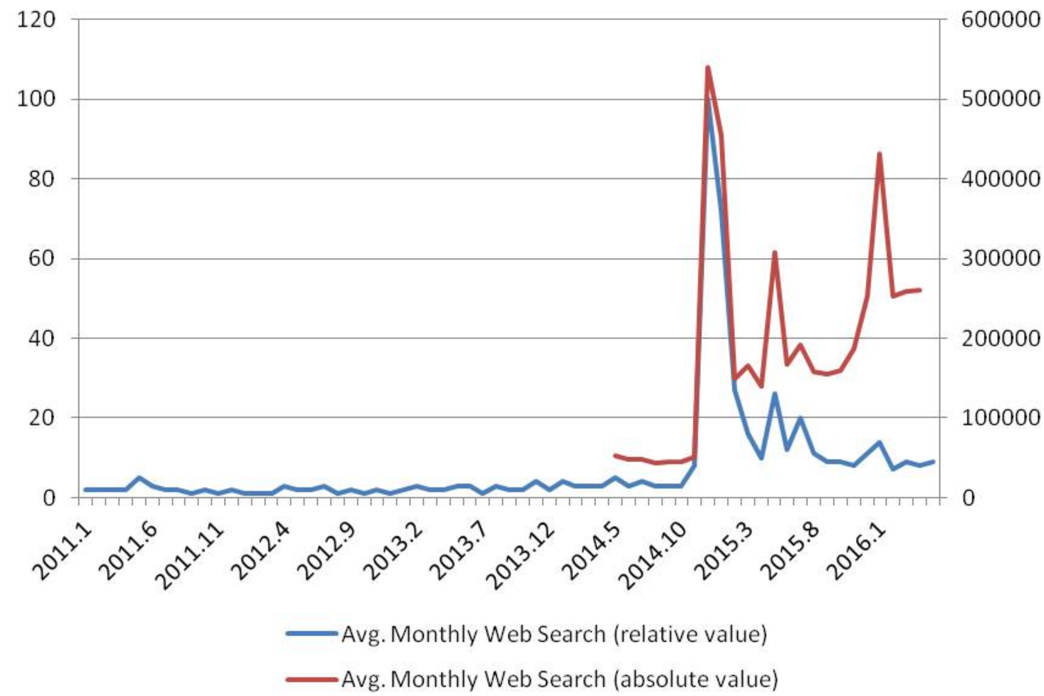
However, at the same time members of Congress have proposed several bills [1] trying to prohibit airlines from penalizing passengers for conducting hidden city ticketing (GAO 2001). In fact, the European Union has passed a passenger bill of rights since around 2005, in which the European Commission has specifically ruled that "airlines must honor any part of an airline ticket" and hidden city ticketing then becomes perfectly legal. After the ruling EU find that "fares have become more fair, hidden city bargains are difficult to find, and the airlines haven't suffered drastic losses due to this".

Although being "threatened" by airline companies, there have been more and more consumers coming to realize the existence of hidden city opportunities, and may try to exploit them to lower their flying costs. In December 2014, United Airlines and Orbitz (an airline booking platform) sued the founder of Skiplagged (a travel search tool) for his website of "helping travelers find cheap tickets through hidden city ticketing". According to CNNMoney, Orbitz eventually settled out of court one year later, and a

[1] According to GAO(2001), several bills proposed in the Congress, including H.R. 700, H.R. 2200, H.R. 5347 and S. 2891, H.R. 332, H.R. 384, H.R. 907 and H.R. 1074, have included language that would not allow airlines to prohibit passengers from hidden-city ticketing.

Chicago judge threw out United's lawsuit using the excuse that the founder "didn't live or do business in that city". And in contrast to the willingness of United, this lawsuit brought the search of key words "hidden city ticketing" to a peak (Figure 2[2]). Corresponding to this higher demand, nowadays there are more travel search

Figure 2: Average monthly web search data of hidden city ticketing.



tools specifically designed to achieve this task (Skiplagged, Tripdelta, Fly Shortcut, AirFareIQ, ITA Matrix, etc). And finding hidden city opportunies and exploiting them become much easier today.

This paper aims at providing some plausible explanations for the cause of hidden city ticketing, and estimating its possible impact on welfare outcomes for airlines, consumers, and society as a whole. I build a structural model in which airlines can choose both prices and network structures as their strategic variables following Shy (2001), and derive several propositions based on that. Then I collect daily flights

[2]Data source for the relative value is Google Trends. Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. Data source for the absolute value is Google AdWords, unit is number of times.

information by scraping the Skiplagged to build my own empirical dataset. I apply global optimization algorithms to estimate the parameters of my model, and then conduct counterfactual analysis to evaluate the possible impact of hidden city ticketing on airlines' expected profits, consumers' welfare, and total surplus, based on which I could help shed some light on policy implications.

Literature Review

To the best of my knowledge, this is the first paper to quantitatively study the cause and impact of hidden city ticketing on welfare outcomes using real empirical data. In fact, there are only a few papers paying attention to this phenomenon. One government report from the Government Accounting Office (GAO 2001) conducted some correlation analysis based on their selected data, and found that the possibility of hidden city ticketing is significantly affected by the size of the markets and the degree of competition in the hub markets and the spoke markets. Another report from Hopper Research (Surry 2016) also provides some summary statistics of this phenomenon. Based on four weeks of airfare search data from Hopper, the analyst found that 26% of domestic routes could be substituted by some cheaper options through hidden city ticketing, and the price discount could be nearly 60%. The most quantitative study is Wang and Ye (2015), which applied a network revenue management model to look at the cause and impact of hidden city ticketing. They base all their findings on simulated data rather than real world data. Therefore, their model is quite different from an economic model. They find that hidden city opportunity may arise when the price elasticity of demand on different routes differ a lot. In order to eliminate any hidden city opportunities, airlines will rise the prices of certain itineraries and hurt consumers. But even airlines optimally react, they will still suffer from a loss in revenue.

There have been a lot of literature focusing on the airline industry ever since its deregulation in 1978. A bunch of them have confirmed significant difference of price elasticity lying between tourists and business travelers. For example, Berry and Jia

(2010) has estimated a price elasticity of demand for tourists as -6.55, while that for business travelers is only -0.63. Pindyck and Rubinfeld (2001) also find a large difference between price elasticity of demand for business travelers (-0.9 to -0.3) and that for leisure travelers (about -1.5). Based on these findings, researchers have further found that airlines are exploiting these differences and engaging in second-degree price discrimination through many different methods, such as advanced-purchase discounts (Dana 1998), ticket restrictions such as Saturday-night stayover requirements (Stavins 2001; Giaume and Guillou 2004), refundable and non-refundable tickets (Escobari and Jindapon 2014), intertemporal price discrimination (Liu 2015; Lazarev 2013), and even the day-of-the-week that a ticket is purchased (Puller and Taylor 2012). This paper follows previous findings and assumes that airline companies are price discriminating between leisure travelers and business travelers, with the latter being less price sensitive and valuing time more. My model also follows Shy (2001) book about economics of network industries, assuming that airlines can choose both airfares and network structures. Finally, I find that while informed passengers could possibly enjoy some benefits of hidden city ticketing, uninformed passengers are always bearing the costs, if any. This is similar to the finding of Varian (1980) where the author shows some "detrimental externalities" that uninformed consumers suffer from due to the behavior of informed consumers.

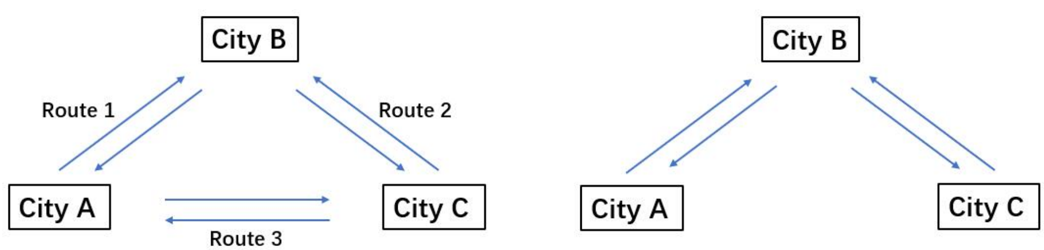
In this paper, I find that 1) hidden city ticketing only occurs when airline companies are applying a hub-and-spoke network structure; 2) under some conditions, hub-and-spoke network is more cost-saving compared to fully connected network; 3) in the short run, hidden city ticketing doet not necessarily decrease airlines' expected profits, while consumers' surplus and total welfare always increase; 4) in the long run, i.e., when airlines can change their choices of prices and networks freely, the impact of hidden city ticketing differs for different routes. For some routes airlines have the incentive to switch from hub-and-spoke network to a fully connected one when there are more and more passengers informed of hidden city ticketing, during which process firms always result in lower expected revenue, while consumers and the whole society are not necessarily better off. Therefore, whether hidden city ticketing should be permitted or forbidden depends on the characteristics of different routes, and this problem cannot be solved by one simple policy.

The remainder of this paper is organized as follows: The structural model is introduced in Section 2, together with the propositions of short run impacts derived from it. Section 3 describes the data in details. Section 4 shows the estimation strategy and the MLE results. In Section 5 I conduct counterfactual analysis to shed some light on long run impacts and policy implications. The limitation and future questions of this paper are discussed in Section 6 and conclude.

2 The Model

Following Shy (2001), I assume that airlines are choosing from two different network structures: fully-connected network or hub-and-spoke network (Figure 3). After the 1978 Airline Deregulation Act, the absence of price and entry controls led to increased use of the hub-and-spoke structure. While in recent years, with the expansion of low-cost carriers (LCCs), fully-connected structure is becoming more popular again. Under

Figure 3: Left: Fully Connected (FC) Network. Right: Hub-and-spoke (HS) Network.



fully-connected network, passengers fly nonstop from one city to the other. While under hub-and-spoke network, everyone who wishes to fly from city A to city C needs to stop at the hub city B. To simplify my analysis below, I use the one-way traveling pattern instead of the two-way traveling pattern showed in Figure 3.

Assume that there is only one airline serving the three cities, thus the firm charges monopoly airfares. Aircrafts are assumed to have an unlimited capacity, thus there is one flight on each route. C_2 denote the airline's cost per mile on any route j. This kind of simplified cost pattern is not uncommon in airline related literature, it is called ACM cost (AirCraft Movement cost) in Shy (2001). Cost pattern can be simplified here mainly because in this specific industry, large percentage of costs are fixed before flights take off (e.g., capital costs, labor costs, etc.) and the marginal cost of airline seats is nearly negligible (Rao 2009). Assume that direct flight has a quality of q_h per mile and indirect flight has a quality of q_l per mile, with $0 < q_l < q_h < 1$.

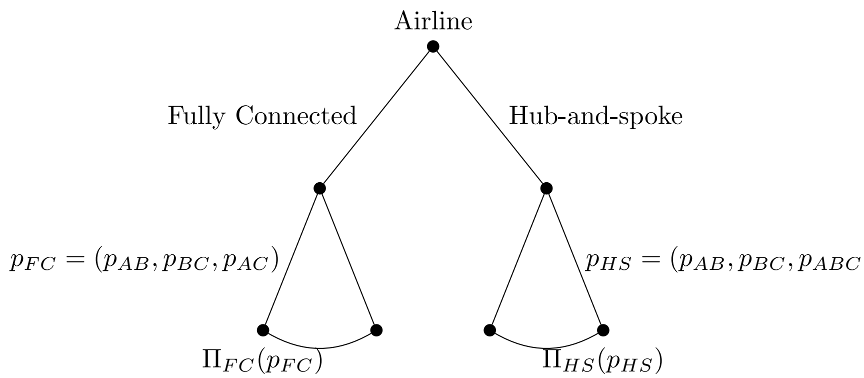
Each individual i has a time preference parameter of λ_i, obtaining utility

$$u_i = C_1 \cdot e^{\lambda_i} q d - p$$

from consuming a good of quality q. And under the assumption of free disposal, he/she will get 0 utility if chooses not to fly. Utility decreases when price increases. If a passenger values time more (i.e., with a larger λ), he/she will enjoy a larger utility increase when switching from an indirect flight (with quality q_l) to a direct flight (with quality q_h). And for a longer itinerary (larger d), the utility improvement from indirect flight to direct flight is also larger. C_1 is a scaling parameter to make the utility comparable to dollar value p.

On each route j, the distribution of consumers' time preferences satisfies $\lambda_{ij} \sim N(\theta_j, \sigma_1^2)$. For passengers flying from A to B, the fraction of passengers being aware of hidden city opportunity is δ and the fraction of uninformed passengers is $1 - \delta$. When hidden city opportunity exists (i.e., $p_{AB} > p_{ABC}$), informed passengers will pay p_{ABC} instead, while uninformed passengers will still pay p_{AB}. The amount of passengers on each route j are normalized to 1. And p_j denote the airfare on route j, d_j denote the distance of route j.

In equilibrium, airline chooses both network structures and prices to maximize



expected profits.

According to the assumptions above, on each route j, for each individual i,

$$u_{ij} = C_1 \cdot e^{\lambda_{ij}} qd - p,\ \ \lambda_{ij} \sim N(\theta_j, \sigma_1^2).$$

Therefore, on each route j, the proportion of consumers that will fly equals to:

$$\begin{aligned} Pr\left[u_{ij} \geq 0\right] &= Pr\left[C_1 \cdot e^{\lambda_{ij}} qd \geq p\right] \\ &= Pr\left[\lambda_{ij} \geq ln\left(\frac{p}{C_1 \cdot qd}\right)\right] \\ &= 1 - \Phi_{\theta_j, \sigma_1^2}\left(ln\left(\frac{p}{C_1 \cdot qd}\right)\right). \end{aligned}$$

2.1 Fully Connected Network

Under fully connected network, airline's expected profits (producer surplus) equal to the revenue it collects minus the costs:

$$\begin{aligned} \Pi_{FC} &= \Pi_{AB} + \Pi_{BC} + \Pi_{AC} \\ &= p_{AB} \cdot \left[1 - \Phi_{\theta_{AB}, \sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\ &+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC}, \sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\ &+ p_{AC} \cdot \left[1 - \Phi_{\theta_{AC}, \sigma_1^2}\left(ln\left(\frac{p_{AC}}{C_1 \cdot q_h d_{AC}}\right)\right)\right] - C_2 \cdot d_{AC}. \end{aligned}$$

Obviously, since under fully connected network structure, the only way to fly "in-directly" from A to C is to take the two direct flights A to B and B to C together, therefore, with $p_{AB} < p_{AB} + p_{BC}$, we can easily derive the following proposition:

Proposition I. *Hidden city opportunity does not exist under fully connected network structure.*

Consumer surplus come from the difference between our willingness to pay and the price we actually being charged, which equal to:

$$\begin{aligned} CS_{FC} &= CS_{AB} + CS_{BC} + CS_{AC} \\ &= \int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{AB}\right) dF(\lambda_i) \\ &+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC} - p_{BC}\right) dF(\lambda_i) \\ &+ \int_{ln\left(\frac{p_{AC}}{C_1 \cdot q_h d_{AC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AC} - p_{AC}\right) dF(\lambda_i). \end{aligned}$$

Adding them together, we get our total surplus under fully connected network:

$$\begin{aligned} TS_{FC} &= PS_{FC} + CS_{FC} \\ &= \int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) - C_2 \cdot d_{AB} \\ &+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC}\right) dF(\lambda_i) - C_2 \cdot d_{BC} \\ &+ \int_{ln\left(\frac{p_{AC}}{C_1 \cdot q_h d_{AC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AC}\right) dF(\lambda_i) - C_2 \cdot d_{AC}. \end{aligned}$$

Since the prices we pay are assumed to be equal to the prices airline receives, the prices get canceled out.

2.2 Hub-and-spoke Network (w/o Hidden City Ticketing)

Under hub-and-spoke network structure, let's first look at the simple case when hidden city opportunity does not exist (i.e., $p_{AB} \leq p_{ABC}$). In this situation, airline's expected profits equal to:

$$\begin{aligned}
\Pi_{HS} &= \Pi_{AB} + \Pi_{BC} + \Pi_{ABC} \\
&= p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\
&+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC},\sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\
&+ p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right].
\end{aligned}$$

Proposition II. *If the cost associated with maintaining route 3 is sufficiently large, then the hub-and-spoke network is more profitable to operate than the fully connected network for the monopoly airline.*

To show the proposition above, let's compare airline's expected profits under these two different networks:

$$\begin{aligned}
\Pi_{HS} - \Pi_{FC} &= p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right] \\
&- p_{AC} \cdot \left[1 - \Phi_{\theta_{AC},\sigma_1^2}\left(ln\left(\frac{p_{AC}}{C_1 \cdot q_h d_{AC}}\right)\right)\right] \\
&+ C_2 \cdot d_{AC}.
\end{aligned}$$

Therefore, if the last term is large enough, hub-and-spoke network is more profitable. This is in accordance with the conclusion drawn by Shy (2001) under a different framework, in which the author found that hub-and-spoke network is cost-saving if the fixed cost is large enough. There is also other literature (Caves, Christensen and Tretheway 1984; Brueckner, Dyer and Spiller 1992; Brueckner and Spiller 1994; Berry, Carnall and Spiller 2006) together with anecdotal evidence from the news showing that airlines are applying hub-and-spoke network in order to lower their costs.

Similarly, consumer surplus equal to:

$$
\begin{aligned}
CS_{HS} &= CS_{AB} + CS_{BC} + CS_{ABC} \\
&= \int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{AB}\right) dF(\lambda_i) \\
&+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC} - p_{BC}\right) dF(\lambda_i) \\
&+ \int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_l d_{ABC} - p_{ABC}\right) dF(\lambda_i).
\end{aligned}
$$

And adding them together, we have total surplus equal to:

$$
\begin{aligned}
TS_{HS} &= PS_{HS} + CS_{HS} \\
&= \int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) - C_2 \cdot d_{AB} \\
&+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC}\right) dF(\lambda_i) - C_2 \cdot d_{BC} \\
&+ \int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_l d_{ABC}\right) dF(\lambda_i).
\end{aligned}
$$

2.3 Hub-and-spoke Network (with Hidden City Ticketing)

Now consider the more complicated case when hidden city opportunity exists (i.e., $p_{AB} > p_{ABC}$). Comparing to the above situation when there is no hidden city ticketing, the only difference lies in the informed passengers who fly directly from A to B.

Therefore, the airline's expected profits equal to:

$$\begin{aligned}
\Pi_{HCT} &= \Pi_{AB} + \Pi_{BC} + \Pi_{ABC} \\
&= (1-\delta) \cdot p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\
&+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC},\sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\
&+ p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right] \\
&+ \delta \cdot p_{ABC} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)\right)\right].
\end{aligned}$$

Proposition III. *When airlines do not alter their choices of prices and network structures, hidden city ticketing does not necessarily decrease airline's expected profits.*

To see why this is true, again, let's compare airline's expected profits with and without hidden city ticketing:

$$\begin{aligned}
\Pi_{HCT} - \Pi_{HS} &= \delta \cdot \left\{p_{ABC} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)\right)\right]\right. \\
&- \left. p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right]\right\}.
\end{aligned}$$

Note that $p_{ABC} < p_{AB}$ while $\left[1-\Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] > \left[1-\Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right]$. Although airline suffers from a loss when passengers are paying a lower price, it also enjoys some gain when this lower price attracts more consumers to take the flight. Therefore, how hidden city ticketing will affect airline's expected profits depends on the relative dominance of the above two inequalities.

Consumer surplus equal to:

$$\begin{aligned}
CS_{HCT} &= CS_{AB} + CS_{BC} + CS_{ABC} \\
&= (1-\delta)\int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{AB}\right) dF(\lambda_i) \\
&+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC} - p_{BC}\right) dF(\lambda_i) \\
&+ \int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_l d_{ABC} - p_{ABC}\right) dF(\lambda_i). \\
&+ \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{ABC}\right) dF(\lambda_i).
\end{aligned}$$

Proposition IV. *When airlines do not alter their choices of prices and network structures, consumers are always better off when hidden city ticketing is allowed.*

Proof:

$$\begin{aligned}
CS_{HCT} - CS_{HS} &= \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{ABC}\right) dF(\lambda_i) \\
&- \delta\int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{AB}\right) dF(\lambda_i) \\
&= \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB} - p_{ABC}\right) dF(\lambda_i) \\
&+ \delta\int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(p_{AB} - p_{ABC}\right) dF(\lambda_i) > 0.
\end{aligned}$$

The increase in consumer surplus come from two different parts. Firstly, the existing passengers are now paying a lower price, which gives them extra utility gain. Secondly, some travelers who will not fly under p_{AB} are now participating in this market activity because of the lower price p_{ABC}, and they also enjoy some utility gain to increase the total consumer surplus.

Adding them together, we have total surplus under hidden city ticketing equal to:

$$\begin{aligned} TS_{HCT} &= (1-\delta)\int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) - C_2 \cdot d_{AB} \\ &+ \int_{ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{BC}\right) dF(\lambda_i) - C_2 \cdot d_{BC} \\ &+ \int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_l d_{ABC}\right) dF(\lambda_i) \\ &+ \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i). \end{aligned}$$

Proposition V. *When airlines do not alter their choices of prices and network structures, total social welfare always increase when hidden city ticketing is allowed.*

Proof:

$$\begin{aligned} TS_{HCT} - TS_{HS} &= \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) \\ &- \delta\int_{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)}^{\infty} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) \\ &= \delta\int_{ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)}^{ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)} \left(C_1 \cdot e^{\lambda_i} q_h d_{AB}\right) dF(\lambda_i) > 0. \end{aligned}$$

Total surplus increase because compared to the original price p_{AB}, now there are more passengers choosing to take the flight under the lower price p_{ABC}, and those extra passengers enjoy extra gain. Since I've assumed unlimited capacity for the aircrafts, the whole society can benefit from this change.

Note that Propositions III, IV and V are derived under the assumption that airlines are not aware of hidden city ticketing, thus they do not alter their choices of prices and network structures in reaction to this booking strategy. In such a case, it can be showed that the model is solvable and the solution is unique (more details in the

Appendix). However, this might be valid in the short run, while in a longer horizon, airlines should be able to realize the behavior of hidden city ticketing and adjust their prices and networks in response to that in order to maximize their revenue. In such a case, obtaining a closed-form solution is difficult, and to see what would be airline's optimal joint choice of prices and network structures when δ changes is even more challenging. Therefore, in the following sections I'm going to use numerical approach instead to solve for the optimal prices and networks under different δ, and estimate the possible impacts of hidden city ticketing on welfare outcomes in the counterfactual analysis below.

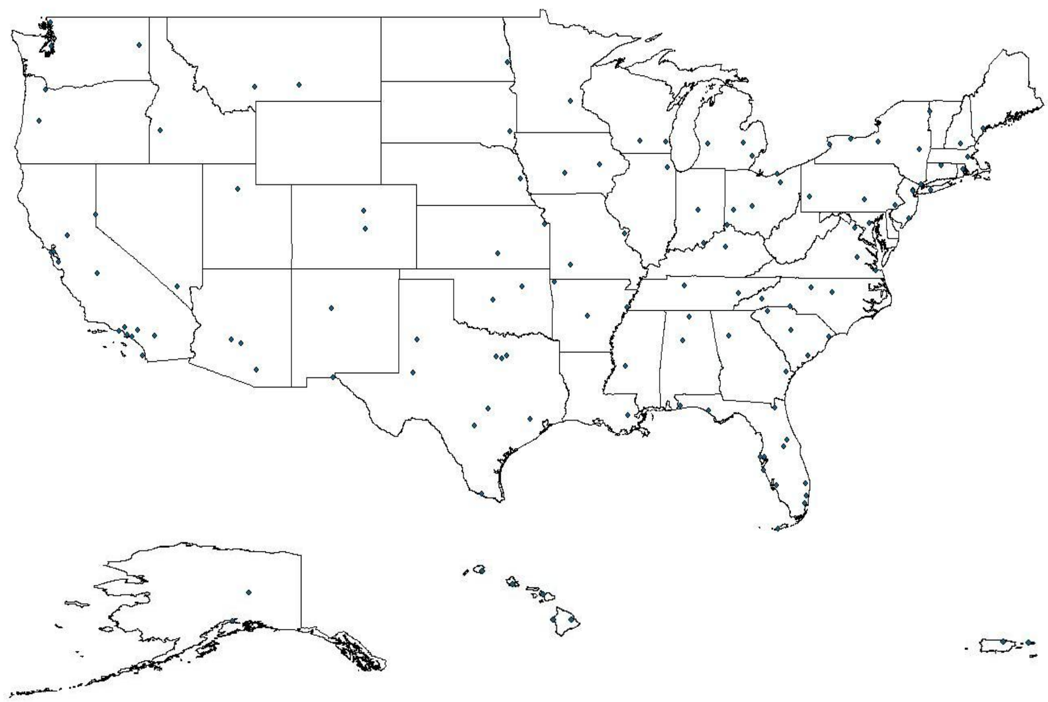
3 Data

I've collected daily flights data by scraping the tickets information on Skiplagged webpage on February 6, 2016 with all quotes of April 6, 2016. This date was chosen because it was neither a weekend nor a holiday, and it was 60 days before the departure date, which should not be severely affected by seat sales. Information being collected include the origin, connecting (if any) and destination airports, time of departure, connection and landing, operation airlines and airfares.

According to the Passenger Boardings at Commercial Service Airports of Year 2014 released in September 2015 by Federal Aviation Administration (FAA), there are more than 500 commercial service airports around United States. To reduce the computational burden of collecting data, I've restricted my sample to the 133 busy commercial service airports identified by FAA. Only focusing on those 133 airports is reasonable because those airports actually accounted for 96.34% of total passenger enplanements in 2014. The distribution of the busy commercial service airports around United States is showed in Figure 4. From the figure we can see that my data has covered airports in Alaska, Hawaii and Puerto Rico, while no airport in Wyoming has been identified as busy commercial service airport in my analysis.

Overall, my sample includes 16,142 routes (airport A to airport B) and 2,822,086

Figure 4: Distribution of busy commercial service airports around United States.



itineraries (flight from A to B with specific information of time, connection node, operation airline(s) and airfare(s)). Flights are operated by 45 different airline companies, among which 11 companies had some hidden city opportunities lying in the itineraries they operated.

To the best of my knowledge, there is no official definition of hidden city opportunity. In GAO (2001), the authors defined "a hidden-city ticketing opportunity to exist for business travelers if the difference in airfares between the hub market and the spoke airport was $100 or more, and for leisure passengers if the difference in airfares was $50 or more". In this paper, I have constructed two intuitive definitions of hidden city opportunity myself below.

- **Definition #1:** Hidden city opportunity exists if the cheapest non-stop ticket of that itinerary is still more expensive than some indirect flight ticket with the direct destination as a connection node.

- **Definition #2:** Hidden city opportunity exists if the non-stop ticket is more expensive than some indirect flight ticket which shares exactly the same first segment of that itinerary.

The second definition is the same as the one being used in Wang and Ye (2015), and is also the example being illustrated at the beginning of the introduction.

According to the daily flights data I have collected, these two definitions show similar magnitude of hidden city opportunities. Among all the itineraries, the first definition indicates a total of 366,754 (13.00%) flights and 1,095 (6.78%) routes which exhibit possible hidden city opportunity. And the numbers from definition #2 are 394,544 (13.98%) and 1,316 (8.15%) respectively. This magnitude is a little bit smaller compared to GAO (2001), in which the authors found that among the selected markets for six major U.S. passenger airlines in their data, 17% provided such opportunities. Table 1 shows the top 10 origin-destination pairs with most hidden city opportunities, which are the same under both definitions.

Table 1: Top 10 Origin-Destination pairs with most hidden city itineraries

Popularity	Origin	Destination	# of Itineraries	% under Def.1	% under Def.2
1	ISP	PHL	11105	3.03%	2.81%
2	SRQ	CLT	8590	2.34%	2.18%
3	CAK	CLT	5948	1.62%	1.51%
4	GRR	ORD	5733	1.56%	1.45%
5	MSN	ORD	5640	1.54%	1.43%
6	XNA	ORD	5529	1.51%	1.40%
7	COS	DEN	4000	1.09%	1.01%
8	FSD	ORD	3665	1.00%	0.93%
9	CAE	CLT	3659	1.00%	0.93%
10	ORF	CLT	3540	0.97%	0.90%

Furthermore, the maximum payment reduction would be as large as 89.57% if hidden city ticketing is allowed. Table 2 shows the top 10 origin-destination pairs with the largest price differences, which are slightly different under both definitions. These findings help shed some light on the fact that hidden city ticketing might no longer be negligible nowadays and related research becomes necessary and valuable.

Table 2: Top 10 Origin-Destination pairs with largest price differences

Definition #1			Definition #2		
Origin	Destination	% Saving	Origin	Destination	% Saving
LGA	IAH	89.57%	LGA	IAH	89.57%
CLE	IAH	88.49%	CLE	IAH	88.49%
PHL	DTW	87.54%	PHL	DTW	87.54%
IAH	EWR	86.61%	MKE	MSP	86.65%
IAH	IAD	86.36%	IAH	EWR	86.61%
DTW	PHL	86.20%	IAH	IAD	86.36%
KOA	SFO	85.87%	DTW	PHL	86.20%
SNA	SLC	85.46%	KOA	SFO	85.87%
ICT	MSP	85.38%	SNA	SLC	85.46%
CLE	EWR	85.03%	ICT	MSP	85.38%

Recall that my primary data contains flights operated by 45 different airline companies, among which 11 companies had some hidden city opportunities lying in the itineraries they operated. Table 3 exhibits the amounts of hidden city itineraries of these airlines unber both definitions. We can see that the three largest airlines: American Airlines, Delta Air Lines and United Airlines operated more than 99% of those itineraries. This is similar to the findings of Surry (2016), in which he found that 96% of those hidden city discounts came from American Airlines, Delta Air Lines, United Airlines and Alaska Airlines. All of them are major hub-and-spoke carriers and apply a hub-and-spoke network business model.

Table 3: Number of hidden city itineraries of different airlines

Airline	IATA Code	Def.1: # of Itineraries (%)	Def.2: # of Itineraries (%)
American Airlines	AA	203096 (55.38%)	210287 (53.30%)
Delta Air Lines	DL	93062 (25.38%)	106867 (27.09%)
United Airlines	UA	69587 (18.98%)	76175 (19.31%)
Alaska Airlines	AS	598 (0.16%)	666 (0.17%)
Hawaiian Airlines	HA	221 (0.06%)	221 (0.06%)
Frontier Airlines	F9	56 (0.02%)	157 (0.04%)
JetBlue Airways	B6	48 (0.01%)	106 (0.03%)
Virgin America	VX	29 (0.01%)	36 (0.01%)
Silver Airways	3M	11 (0.00%)	11 (0.00%)
Spirit Airlines	NK	8 (0.00%)	11 (0.00%)
Sun Country Airlines	SY	7 (0.00%)	7 (0.00%)

A notable exception is Southwest Airlines, where no hidden city opportunity is found in the itineraries operated by it, and whose fare rules actually do not specifically prohibit the practice of hidden city ticketing. Since Southwest Airlines applies fully connected network, this finding in the data is in accordance with my previous proposition that hidden city opportunity does not exist under fully connected network structure.

4 Estimation

To estimate the parameters of my model, firstly I retrieve all the ordered triplets (A-B-C) from my primary dataset. Then, with all the observed information of prices, distances and consumers' preferences, I choose the parameters of my model to maximize the likelihood of observed airlines' choices of network structures. In order to deal with this implicit maximum likelihood function, I apply global optimization algorithms, more specifically, Pattern Search to estimate the parameters.

4.1 Sample Build-up

I've retrieved ordered triplets (A-B-C) from the 133 busy commercial service airports of my primary dataset. The triplet needs to satisfy following conditions: 1) it includes direct flight from A to B; 2) it includes direct flight from B to C; 3) it includes either direct or one-stop indirect flight from A to C using B as the connection node.

I've obtained 114,635 ordered triplets from my dataset. Based on the differences in condition #3, I divide them into three different types. Type I includes only direct flight from A to C with a subsample size of 26,198. To estimate p_{ABC} of Type I, I add observed p_{AB} and p_{BC} up manually. Type II includes only indirect flight from A to C through B with a subsample size of 61,092. To estimate p_{AC} of Type II, I use the observed p_{AB} and assume flights from A to B and A to C share the same price per mile: $p_{AC} = \frac{p_{AB}}{d_{AB}} \cdot d_{AC}$. Type III includes both direct flight from A to C and indirect flight from A to C through B, with a subsample size of 27,345. d represents geodesic distance being computed using the longitude and latitude of the pair-airports provided by Google Maps.

On each route j, assume that $\theta_j \sim N(\mu_j, \sigma_2^2)$. Recall that each individual i has a time preference parameter of λ_i and on each route j, the distribution of consumers' time preferences satisfies $\lambda_{ij} \sim N(\theta_j, \sigma_1^2)$. Therefore, μ_j measures the dependency of the destination city on business travelers. Previous literature have constructed several indexes to capture this characteristic. For example, Borenstein (1989) and Borenstein and Rose (1994) built a tourism index at the MSA level based on the ratio of hotel income to total personal income. Brueckner, Dyer and Spiller (1992) and Stavins (2001) assumed that the difference in January temperature between origin and destination cities could serve as a proxy for tourism. And Gerardi and Shapiro (2009) segmented their data into "leisure routes" and "big-city routes" based on the ratio of accommodation earnings to total nonfarm earnings.

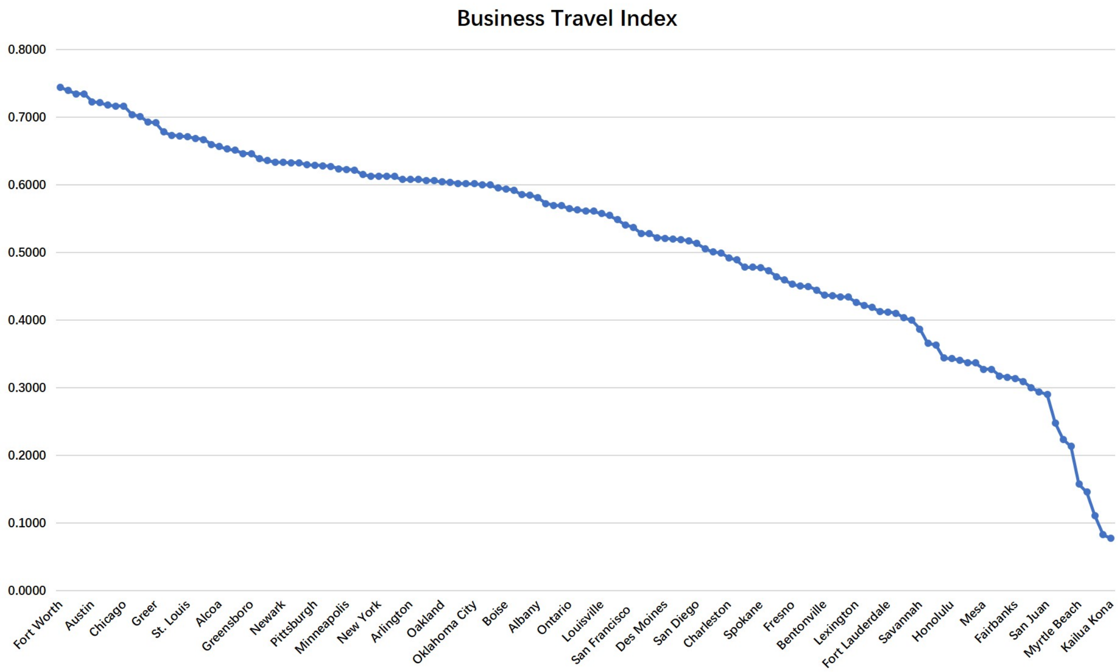
In this paper, I've constructed my own index according to Borenstein (2010) and data provided by TripAdvisor. Borenstein (2010) provides an index of the share of commercial airline travel to and from cities that is for business purposes, which is based on the 1995 American Travel Survey. This index was also used as one of the measures in Puller and Taylor (2012) to distinguish between "leisure" and "mixed" routes. The shortage for this index is that it only includes data for each state and MSA, while city level data might be a better fit corresponding to the location of an airport. To deal with this issue, I've also collected data from TripAdvisor (https://www.tripadvisor.com/ the largest travel site in the world) for each city, and compute the average number of the reviews of hotels/lodging, vocation rentals, things to do, restaurants, and posts of forum, standardized by the city population from 2010 census. The underlying assumption is that a larger number of reviews on TripAdvisor might be an indicator of being more popular among leisure travelers, and this city-level data together with the indices constructed by Borenstein (2010) might be able to provide more complete information of the city's characteristics. After taking exponential of the opposite of the average number from TripAdvisor's review data, I compute the mean of that and the indices from Borenstein (2010) (both state-level and MSA-level) and get μ.

From Figure 5 we can see that the largest $\mu = 0.7450$ goes to Dallas Fort Worth International Airport (DFW) in Texas, while the smallest $\mu = 0.0773$ goes to Ellison Onizuka Kona International Airport (KOA) on the Island of Hawaii. In general, places that are more popular among tourists, such as Orlando, Puerto Rico and Hawaii, get the smaller μ(s). While places like Dallas, Austin and Chicago that are more attractive to business travelers have larger μ(s).

4.2 Maximum Likelihood Estimation

Overall, I have a set of 7 parameters: $\zeta = (\delta, C_1, C_2, \sigma_1, \sigma_2, q_h, q_l)$, with $\delta \in [0, 1]$ as my parameter of interest, and the others are nuisance parameters. Observed attributes in

Figure 5: Business travel index for each airport as the destination city.



my dataset include the prices, distances, and time preference indices on each route: $x_i = (p_{AB,BC,AC,ABC}, d_{AB,BC,AC,ABC}, \mu_{AB,BC,AC,ABC})$. And observed decision variable is the airline's network choices: $y_i \in \{FC, HS\}$.

The maximum likelihood estimation needs to be processed in 2 steps. Firstly, I sample $\theta_{AB,BC,AC,ABC}$ from the normal distribution $\mathbb{N}|_{x_i,\sigma_2}$. Then airline makes a decision to maximize expected profits:

$$y_i = \arg \max_{y \in \{FC,HS\}} \Pi(x_i, y, \zeta).$$

The maximum likelihood estimation problem is therefore:

$$\hat{\zeta} = \arg\max_{\zeta} \frac{1}{n} \sum_{i=1}^{n} \log\, p(y_i | x_i; \zeta),$$

with the probabilistic model as

$$Pr[y_i = y|x_i, \zeta] = \underset{\theta \sim N|_{x_i,\sigma_2}}{Pr} \left[\Pi(x_i, y, \zeta, \theta) \geq \Pi(x_i, \neg y, \zeta, \theta)\right].$$

4.3 Pattern Search

This maximum likelihood estimation is challenging because the likelihood is implicit with a random sampling in the first step, and the gradient is also difficult to evaluate with respect to ζ. Here I apply global optimization algorithms to maximize this log likelihood. That is, for each ζ_t, obtain an estimation of likelihood function

$$\begin{aligned} \log p(y_i|x_i; \zeta_t) &= \log \underset{\theta \sim \mathbb{N}|_{x_i,\sigma_2}}{Pr} \left[\Pi(x_i, y_i, \zeta_t, \theta) \geq \Pi(x_i, \neg y_i, \zeta_t, \theta)\right] \\ &\approx \log \left\{ \frac{1}{M} \sum_{m=1}^{M} \mathbf{1} \left[\Pi(x_i, y_i, \zeta_t, \theta_m) \geq \Pi(x_i, \neg y_i, \zeta_t, \theta_m)\right] \right\}. \end{aligned}$$

I've tried several global optimization techniques including Pattern Search, Genetic Algorithm, Simulated Annealing, etc., to get the optimal $\zeta = (\delta, C_1, C_2, \sigma_1, \sigma_2, q_h, q_l)$ that maximizes the log likelihood. It turns out that Pattern Search works best in this case. It costs the shortest time; It achieves the maximum log likelihood; And it obtains quite similar and robust results when I change the starting point from $\delta = 0.1, 0.5$ to 0.9.

Pattern Search algorithm fits this problem quite well because firstly, it does not require the calculation of gradients of the objective function, which are quite difficult to compute in this case. Secondly, it lends itself to constraints and boundaries. For example, it could deal with the constraint that $0 < q_l < q_h < 1$ quite well in this case.

How the Pattern Search algorithm operates? Pattern search applies polling method (Mathworks 2018) to find out the minimum of the objective function. Starting from an initial point, it firstly generates a pattern of points, typically plus and minus the coordinate directions, times a mesh size, and center this pattern on the current point. Then, for each point in this pattern, evaluate the objective function and compare to the evaluation of the current point. If the minimum objective in the pattern is smaller than the value at the current point, the poll is successful, and the minimum point found becomes the current point. The mesh size is then doubled in order to escape from a local minimum. If the poll is not successful, the current point is retained, and the mesh size is then halved until it falls below a threshold when the iterations stop. Multiple starting points could be used to insure that a robust minimum point has been reached regardless of the choice of the initial point.

This algorithm is simple but powerful, provides a robust and straightforward method for global optimization. It works well for the maximum likelihood function in this paper, which is derivative-free with constraints and boundaries.

4.4 Estimation Results

The estimation results from Pattern Search are shown in Table 4 below.

Table 4: Results of MLE

log likelihood	-0.3023
δ	0.0373
C_1	10.0935
C_2	0.3125
σ_1	0.2094
σ_2	0.7406
q_h	0.7010
q_l	0.1125

According to the estimation results, the informed passengers account for around 3.73% of the whole population. This number seems to be small at first glance, but it is actually not surprising because those are the travelers who are not only informed about hidden city ticketing, but also exploiting those opportunities and resulting in affecting the choices made by airlines. And in the counterfactual analysis section below, you are going to see that even a small fraction of informed passengers will affect airline's choices of network structure and prices significantly.

In order to derive confidence interval of my parameter of interest δ, again I apply numerical approach using bootstrap to find out standard errors. I run the MLE for 1000 times, and for each run, sample the entire data with replacement and construct a data set of equal size. The sample mean of the 1000 estimates of the MLE is 0.0296, and the standard error is the sample standard deviation as 0.0093. We can see that δ is significantly different from zero at 99% confidence interval.

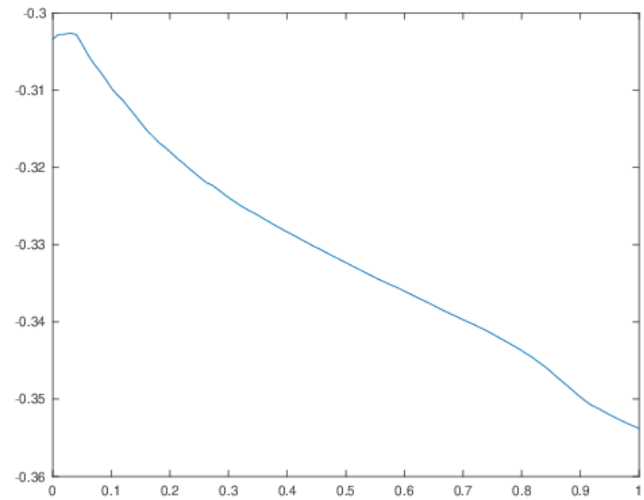
In Figure 6 below, I've plotted the log likelihood as a function of δ with all other parameters being constant at their optimal values. From the figure it is clear that $\delta = 3.73\%$ is the global maximizer.

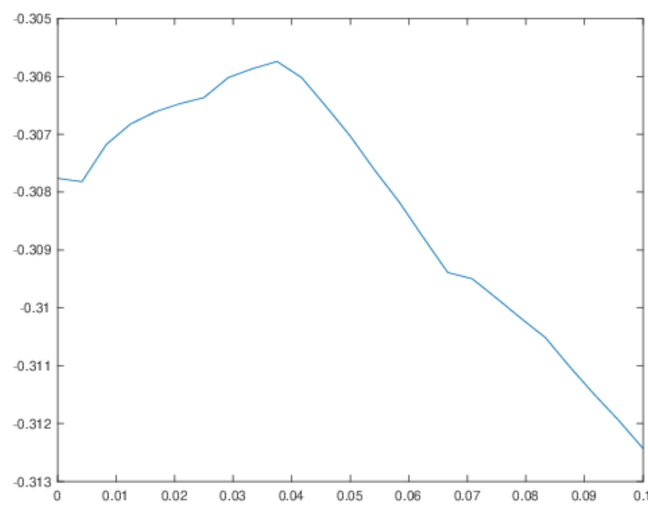
5 Counterfactual Analysis

As I've mentioned before, given a longer horizon, airlines should be able to adjust their prices and networks in response to hidden city ticketing to maximize their revenue. To see what would be airline's optimal joint choice of prices and network structures when δ changes, and further estimate the possible impacts of hidden city ticketing on welfare outcomes, I've conducted several counterfactual analysis using numerical approach below. Will hidden city ticketing always be good for consumers and the whole society? Should government make policies to clearly forbid or permit this booking ploy? My counterfactual experiments are going to help shed some light on those important policy implications. And you are going to see that things are quite complicated in reality and it is difficult to make a decision that is always good for everyone.

Basically, let's assume that the proportion of informed passengers (δ) increases from 0 to 100%, and airline companies then choose optimal prices under different network structures to maximize their expected profits according to the game. After obtaining p^* under different networks, I compute the surplus according to the model, and plot producer surplus (blue), consumer surplus (red) and total surplus (black) under fully connected network (dotted line) and hub-and-spoke network (solid line) respectively when δ varies.

Figure 6: Up: Plot of log likelihood when δ varies from 0 to 1. Below: Plot of log likelihood when δ varies from 0 to 0.1 (zoom in).





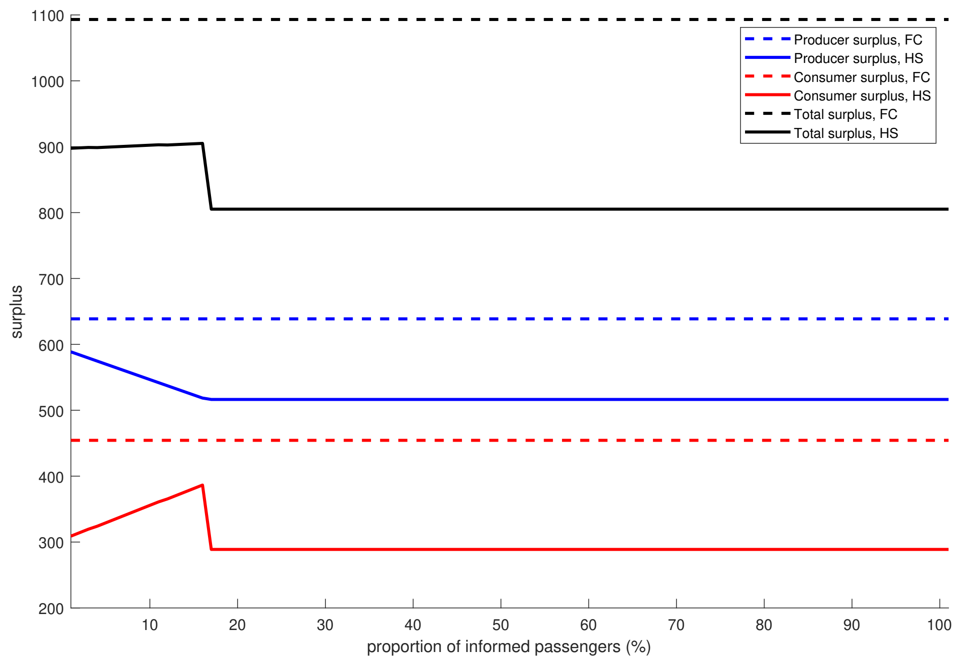
5.1 Fully Connected Network Outperforms Hub-and-spoke Network

Findings I. *Among all the 114,635 data points (i.e., ordered triplets A-B-C), 75,995 (66.29%) have expected profits under fully connected network being always higher than that under hub-and-spoke network, regardless of the value of* δ.

My first finding is that under major cases fully connected network creates higher expected profits for airlines comparing to hub-and-spoke network, regardless of the proportion of informed passengers. One example might be the ordered triplets MIA→SEA→COS (Miami International Airport to Seattle-Tacoma International Airport to Colorado Springs Airport). Figure 7 below shows the surplus under different δ(s). The dotted lines are always horizonal because according to my model, hidden city ticketing will not affect the outcomes under fully connected network structure. It is clear that in this example the dotted blue line is always above the solid one. This should not be surprising because flying from Miami to Colorado through Seattle doesn't make any sense.

When I dig into the details, what happens at the kink is that airline keeps raising the price p_{ABC} in response to the increasing δ. Consumers benefit at first because more and more informed passengers are able to exploit hidden city opportunities and earn some utility gain. However, when the kink point is reached, p_{ABC} hits the magnitude of p_{AB} and hidden city opportunities disappear. Furthermore, since the new p_{ABC} is higher compared to the original one without hidden city ticketing, even those uninformed passengers are getting hurt. This is similar to what was called "detrimental externalities" in Varian (1980), in which the author also found that sometimes more informed consumers would cause the price paid by uninformed consumers to increase. This kind of "kink pattern" is not rare in my data, and it helps confirm the concern mentioned in GAO (2001) that allowing hidden city ticketing might lead to unintended consequences, including higher prices. This is also in accordance with the anecdotal evidence that airlines claim that with more and more passengers conducting hidden

Figure 7: Surplus for MIA→SEA→COS when δ changes.



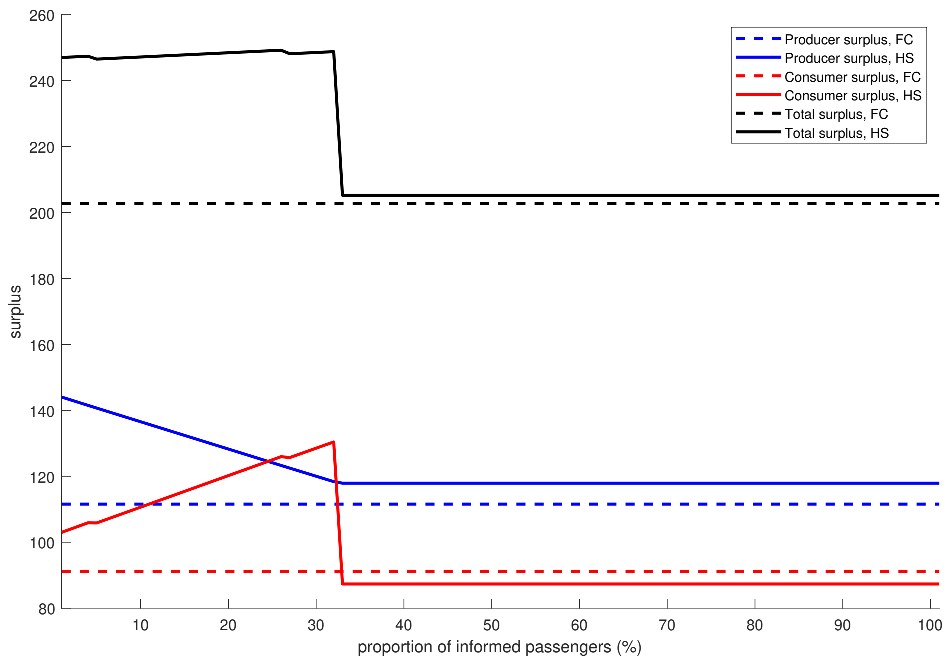
city ticketing, they are going to raise the airfares to eliminate those opportunities and consumers would then be hurt by the higher prices in the long run.

5.2 Hub-and-spoke Network Outperforms Fully Connected Network

Findings II. *22,551 (19.67%) data points have expected profits under hub-and-spoke network being always higher than that under fully connected network, regardless of the value of* δ.

My second finding is that sometimes, hub-and-spoke network structure always does a better job to achieve higher revenue. One example would be the ordered triplets CID→DTW→MSN (The Eastern Iowa Airport to Detroit Metropolitan Airport to Dane County Regional Airport in Madison). Figure 8 below shows the surplus under different δ(s).

Figure 8: Surplus for CID→DTW→MSN when δ changes.



We can see that the solid blue line is always above the dotted one. This kind of pattern usually happens when both airports A and C are small, which is exactly what occurs when you are flying from CID to MSN. In this case, it might be costly for airlines to provide a direct flight service, especially when compared to the relatively low demand.

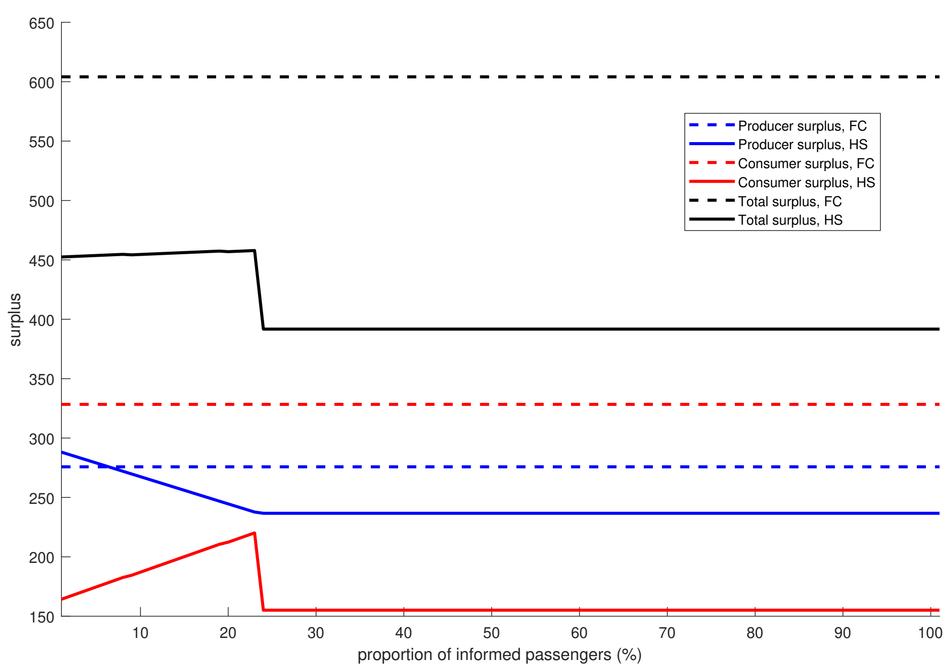
5.3 Switch from Hub-and-spoke Network to Fully Connected Network

Findings III. *16,089 (14.03%) data points have crossings, which means that airline's expected profits are higher under hub-and-spoke network when there are less informed passengers, while fully connected network becomes more profitable when δ is large.*

A more interesting story lies in the cases remained: hub-and-spoke network works well when δ is small, but being outperformed by fully connected network with more and more informed passengers. That is to say, in some cases, airlines have the incentive to switch from one network structure to another, and δ is going to affect companies' network choices. This finding could be supported by what we called "dehubbing" phenomenon in recent years (Berry, Carnall and Spiller 2006). For example, Delta closed its Dallas-Fort Worth International Airport (DFW) hub in 2005 and reduced the number of flights at its Cincinnati hub by 26% in the same year. And Pittsburgh was also downgraded from a hub to a "focus city" by US Airways in 2004.

One example would be the ordered triplets AUS→JFK→RDU (Austin–Bergstrom International Airport to JFK to Raleigh–Durham International Airport). Figure 9 below shows the surplus under different δ(s).

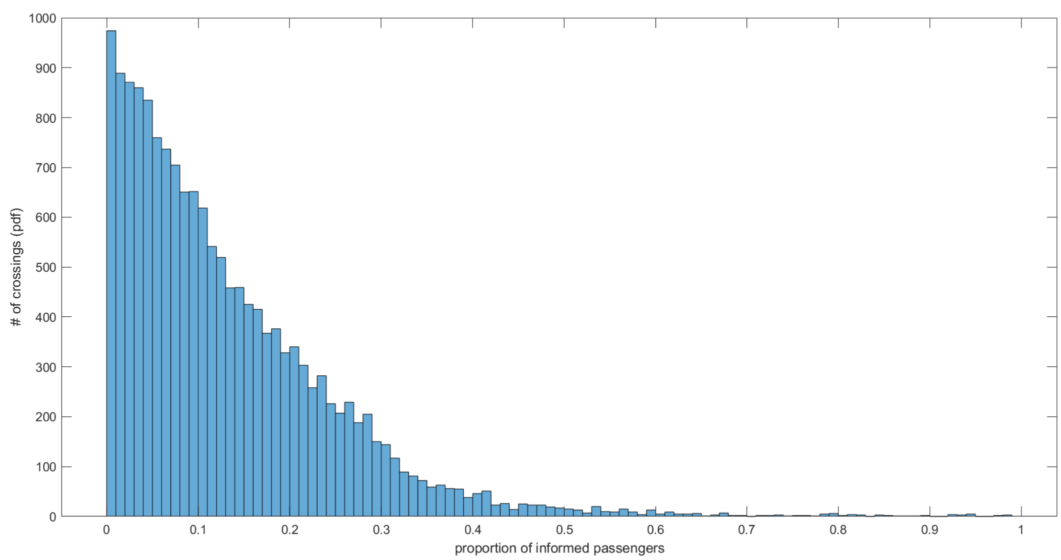
Figure 9: Surplus for AUS→JFK→RDU when δ changes.



We can see that the solid blue line crosses the dotted one at the point when δ is around 6%. Which means that when δ is smaller than that threshhold, airline should pursue hub-and-spoke network structure. While when δ crosses that point, airline has the incentive to switch to another network choice, and this decision is going to affect both consumer surplus and total surplus dramatically. (In this example, both surplus increase a lot when switching from the solid lines to the dotted ones, but this is not always the case.)

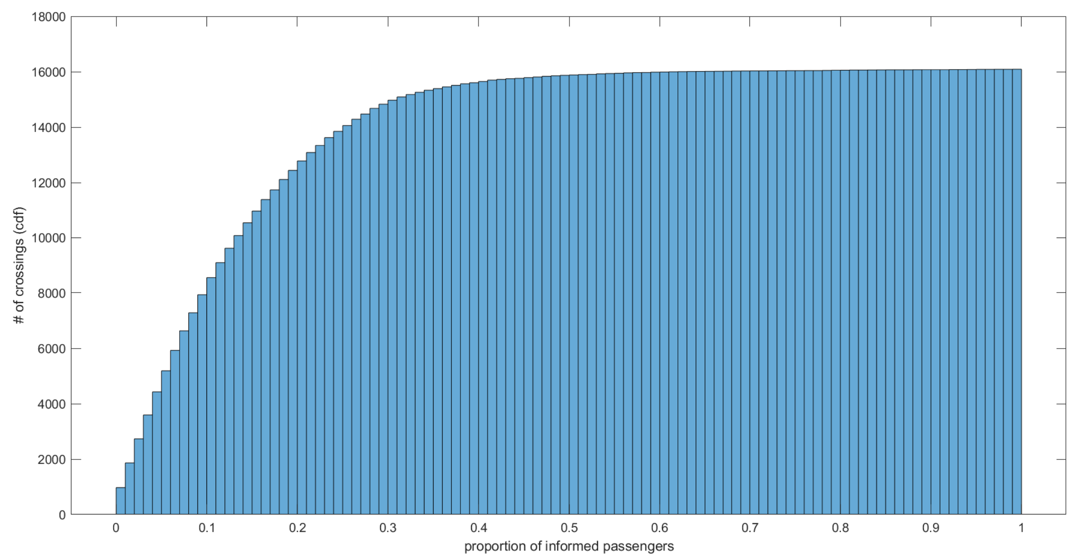
The crossing point varies for different ordered triplets. This is because different routes have different characteristics. Some would be quite "sensitive" to hidden city ticketing and switch when δ is small; some would switch only when informed passengers are more enough; and we've already known that some would never switch; while some would stick to the fully connected network from the very beginning. To have a more complete idea about the impact of different δ on network choices, I'm able to draw the graph for every one data point. Obviously, it is impossible to display those more than one hundred thousands figures here. Therefore in Figure 10 below I've plotted the distribution of all the crossing points when δ changes.

Figure 10: Distribution of crossings when δ changes (pmf)



We can see that even a small δ matters. Airline's choices can be affected significantly even with a quite small proportion of informed passengers. To make the illustration more clear, I've also drawn the cdf of the crossing points, and obtain my next finding from Figure 11 below.

Figure 11: Distribution of crossings when δ changes (cdf)



Findings IV. *Airlines have the incentive to switch from hub-and-spoke network to fully connected network for half of the routes when there are around 10% informed passengers, and for 75% of the routes when δ is only around 19%.*

After comparing the consumer surplus and total surplus before and after switching, I'm able to further conclude that:

Findings V. *If airlines switch from hub-and-spoke network to fully connected network, under 11,458 cases (71.22%) consumer surplus is going to increase, and under 11,128 (69.17%) cases total surplus is going to increase.*

That is to say, unlike what happens under partial equilibrium (or in the short run) when airlines do not alter their choices of network structure and airfares, in the long run, hidden city ticketing might be good for consumers and social welfare in major cases, but not always. There is possibility that consumer surplus and total surplus are getting smaller due to this conduct. Therefore, asking government to make a single policy about permitting or forbidding hidden city ticketing would be tough.

6 Discussions and Conclusion

There are some limitations and future questions remained in this paper. Firstly, I've made a critical assumption in my model that the airline industry is monopoly. This is not uncommon in airline related literature, but some researchers believe that after the deregulation, the US airline industry should be characterized as being hightly oligopolistic (Shy 2001). And according to GAO (2001), "hidden city opportunities may arise when a greater amount of competition exists for travel between spoke communities than on routes to and from hub communities, and where airfares in those markets reflect such competition". Therefore, besides the factors I've raised in my propositions, competition might be another possible cause of hidden city ticketing, which does not enter my model under the monopoly assumption. And given competition, besides the cost-saving consideration, another advantage of hub-and-spoke network compared to fully connected network would be that airlines could have stronger market power in the hub, which helps them increase the entry barrier and drive up prices there. (Borenstein 1989, 1991)

Furthermore, in my model I also assume that airline companies are choosing between fully connected and hub-and-spoke network structures, but I do not provide an outside option for airlines to stop offering the flights for certain routes. For example, instead of raising the price of the indirect flight to eliminate hidden city ticketing, another more extreme possibility would be that the airline simply stop serving certain routes. This is supported by some anecdotal evidence, for example, in some news airlines "threatened" that they would stop service if they caught more and more hidden city ticketing. And this is also another major concern in GAO (2001) that allowing hidden city ticketing might result in unintended decreasing service. This impact is difficult to evaluate without a good measure of the costs, but we could possibly think of the results from this paper as more likely to be a lower bound of the effects. A similar concern would be that airlines have to re-calibrate their no show algorithms and react to the change by adjusting overbooking decisions, which might be difficult and expensive.

Another question raised by Varian (1980) would be that what if it pays to be informed. This is definitely the case here because passengers need to "bear some risk" to conduct hidden city ticketing. On one side, as I've quoted from the contract of carriage, cost of conducting hidden city ticketing is not zero, consumers will be penalized if being caught. And airlines keep emphasizing that hidden city ticketing is "unethical" and doing so "is tantamount to switching price tags to obtain a lower price on goods sold at department stores". It causes delays and sometimes they need to double check checked baggage. On the other side, there are many situations that hidden city ticketing is restricted. For example, if you have luggage that is not carry-on, you can not leave the flight earlier without picking up your bag: normally your checked baggage won't be delivered to your connection city. And you cannot conduct hidden city ticketing for the first segment of your round-trip: your second trip will be cancelled if you missed the connection of the first one. You might also need to bear the risk that you are switched to another flight with the same origin and destination airports, but a different connection city. Therefore, assume some cost will incur to be informed might be a reasonable setting in future research, while measuring this cost might also be challenging.

To conclude, this paper aims at analyzing the possible cause and impact of hidden city ticketing. To achieve this goal, I've constructed a structural model, collected innovative data, applied estimation algorithm and conducted several counterfactual analysis. I've found that hidden city ticketing occurs only when airline company is applying hub-and-spoke network structure. And airline applies hub-and-spoke network other than fully connected network in order to reduce costs. Under partial equilibrium (when airlines do not alter their choices of prices and network structures), according to the theoretical model, 1) hidden city ticketing might increase airline's expected profits when it attracts more passengers to take the flights; 2) consumers are always better off when hidden city ticketing is allowed; 3) total social welfare always increase when hidden city ticketing is allowed.

Things get more complex in the long run. Under general equilibrium (when airlines are free to alter their optimal choices of prices and network structures), according to the counterfactual analysis, 1) to maximize expected profits, fully connected network is always better than hub-and-spoke network for some routes (66.29%), while hub-and-spoke network outperforms fully connected network for some other routes (19.67%), regardless of the proportion of informed passengers; 2) for the rest (14.03%) of the routes, airlines' expected profits are higher under hub-and-spoke network when there are less informed passengers, while fully connected network becomes more profitable when there are more and more passengers exploiting hidden city opportunities; 3) airlines have the incentive to switch from hub-and-spoke network to fully connected network for half of the routes when there are around 10% informed passengers, and for 75% of the routes when informed passengers increase to around 19%; 4) if airlines switch from hub-and-spoke network to fully connected network due to more and more informed passengers, firms are getting worse, while consumers are not always benefited (28.78% cases surplus decrease), and total social welfare is not always larger (30.83% cases surplus decrease). Therefore, asking government to make a single policy about permitting or forbidding hidden city ticketing would be tough.

Appendices

To show that the model is solvable and has unique solution when airlines do not alter their choices of prices and network structures, according to the previous analysis,

$$
\begin{aligned}
\Pi_{FC} &= p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\
&+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC},\sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\
&+ p_{AC} \cdot \left[1 - \Phi_{\theta_{AC},\sigma_1^2}\left(ln\left(\frac{p_{AC}}{C_1 \cdot q_h d_{AC}}\right)\right)\right] - C_2 \cdot d_{AC}.
\end{aligned}
$$

$$
\begin{aligned}
\Pi_{HS} &= p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\
&+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC},\sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\
&+ p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right]
\end{aligned}
$$

if $p_{AB} \leq p_{ABC}$, and

$$
\begin{aligned}
\Pi_{HS} &= (1-\delta) \cdot p_{AB} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}\right)\right)\right] - C_2 \cdot d_{AB} \\
&+ p_{BC} \cdot \left[1 - \Phi_{\theta_{BC},\sigma_1^2}\left(ln\left(\frac{p_{BC}}{C_1 \cdot q_h d_{BC}}\right)\right)\right] - C_2 \cdot d_{BC} \\
&+ p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right] \\
&+ \delta \cdot p_{ABC} \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}\right)\right)\right]
\end{aligned}
$$

if $p_{AB} > p_{ABC}$.

When airlines are not aware of hidden city ticketing thus do not alter their choices of prices and network structures, solving for the optimal price bundle $(p_{AB}, p_{BC}, p_{AC}, p_{ABC})$

to maximize revenue Π_{FC} and Π_{HS} is equivalent to solving the following problem:

$$\max_p \; p \cdot \left[1 - \Phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right)\right]$$

where C is a constant.

Take derivative of the objective function:

$$\begin{aligned} 1 - \Phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) + p \cdot \left[-\phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) \cdot \frac{C}{p} \cdot \frac{1}{C}\right] &= 0 \\ 1 - \Phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) - \phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) &= 0. \end{aligned}$$

Let $x = ln(\frac{p}{C})$, $y = \Phi_{\theta,\sigma^2}(x)$, the equation above becomes a typical ODE:

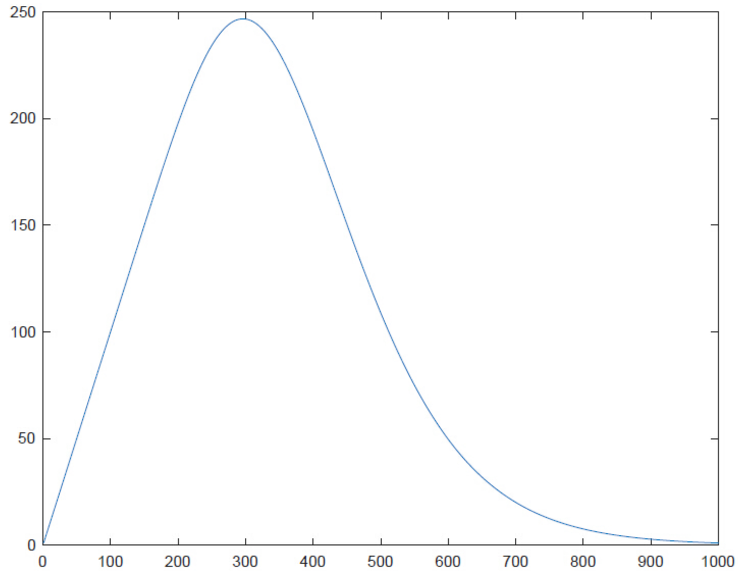
$$\begin{aligned} 1 - y &= \frac{dy}{dx} \\ dx &= \frac{dy}{1-y} \\ x &= -ln(1-y) + C_1 \\ ln(\frac{p}{C}) &= -ln(1-y) + C_1 \\ \frac{p}{C} &= e^{-ln(1-y)} \cdot e^{C_1} \\ &= \frac{e^{C_1}}{1-y} \\ p &= \frac{C \cdot e^{C_1}}{1-y} \\ p - p \cdot \frac{1}{2}\left[1 + \text{erf}\left(\frac{ln(\frac{p}{C}) - \theta}{\sigma\sqrt{2}}\right)\right] &= C \cdot e^{C_1} \end{aligned}$$

where

$$\begin{aligned} \text{erf}(z) &= \frac{1}{\sqrt{\pi}} \int_{-z}^{z} e^{-t^2} dt \\ &= \frac{2}{\sqrt{\pi}}\left(z - \frac{z^3}{3} + \frac{z^5}{10} - \frac{z^7}{42} + \frac{z^9}{216} - \cdots\right) \end{aligned}$$

by Taylor expansion.

To further confirm that function $f(p) = p \cdot \left[1 - \Phi_{\theta,\sigma^2}\left(ln\left(\frac{p}{C \cdot q \cdot d}\right)\right)\right]$ is uni-modal, I depict function $f(p)$ with parameters $\theta = 0.5,\ \sigma = 0.3,\ C = 10,\ q = 0.8,\ d = 30$, as shown in the figure below.



Therefore, the solution of optimal price bundle $(p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ exist and it is unique, under the assumption that airlines do not alter their choices of prices and network structures.

When airlines realize and react to hidden city ticketing, the difference lies in the optimal value of p_{ABC}. Instead of looking for a p_{ABC} that maximizes $p_{ABC} \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(ln\left(\frac{p_{ABC}}{C_1 \cdot q_l d_{ABC}}\right)\right)\right]$, we are now solving the following problem instead:

$$\max_{p}\ p \cdot \left[1 - \Phi_{\theta_1,\sigma^2}\left(ln(\frac{p}{C_1})\right)\right] + \delta \cdot p \cdot \left[1 - \Phi_{\theta_2,\sigma^2}\left(ln(\frac{p}{C_2})\right)\right]$$

where C_1 and C_2 are constants.

Take derivative of the objective function obtains:

$$1-\Phi_{\theta_1,\sigma^2}\left(ln(\frac{p}{C_1})\right)-\phi_{\theta_1,\sigma^2}\left(ln(\frac{p}{C_1})\right)+\delta\left[1 - \Phi_{\theta_2,\sigma^2}\left(ln(\frac{p}{C_2})\right) - \phi_{\theta_2,\sigma^2}\left(ln(\frac{p}{C_2})\right)\right] = 0.$$

Left-hand side is a function of p, $f(p)$ with $p \in [0, +\infty)$. It is continuous because it is a linear combination of probability density function and cumulative distribution function of normal distribution, which are all continuous functions.

When $p \to 0$, $\Phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) \to 0$, $\phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) \in (0,1)$. Therefore, $f(p) > 0$. When $p \to +\infty$, $\Phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) \to 1$, $\phi_{\theta,\sigma^2}\left(ln(\frac{p}{C})\right) \in (0,1)$. Therefore, $f(p) < 0$. According to Mean Value Theorem, there exists at least one p that makes $f(p) = 0$. Therefore, the solution of the model still exists.

However, solving for a closed-form solution is challenging, numerical approach might be a better choice in this case.

References

[1] Berry, S., Carnall, M., & Spiller, P. T. (2006). Airline hubs: costs, markups and the implications of customer heterogeneity. *Competition policy and antitrust.*

[2] Berry, S., & Jia, P. (2010). Tracing the woes: An empirical analysis of the airline industry. *American Economic Journal: Microeconomics*, 2(3), 1-43.

[3] Borenstein, S. (1989). Hubs and high fares: dominance and market power in the US airline industry. *The RAND Journal of Economics*, 344-365.

[4] Borenstein, S. (1991). The dominant-firm advantage in multiproduct industries: Evidence from the US airlines. *The Quarterly Journal of Economics*, 106(4), 1237-1266.

[5] Borenstein, S., & Rose, N. L. (1994). Competition and price dispersion in the US airline industry. *Journal of Political Economy*, 102(4), 653-683.

[6] Borenstein, S. (2010). An index of inter-city business travel for use in domestic airline competition analysis. *UC Berkeley, Index available at http://www.nber.org/data/bti.html (accessed 08.02.14).*

[7] Brueckner, J. K., Dyer, N. J., & Spiller, P. T. (1992). Fare determination in airline hub-and-spoke networks. *The RAND Journal of Economics*, 309-333.

[8] Brueckner, J. K., & Spiller, P. T. (1994). Economies of traffic density in the deregulated airline industry. *The Journal of Law and Economics*, 37(2), 379-415.

[9] Caves, D. W., Christensen, L. R., & Tretheway, M. W. (1984). Economies of density versus economies of scale: why trunk and local service airline costs differ. *The RAND Journal of Economics*, 471-489.

[10] Dana, Jr, J. D. (1998). Advance-purchase discounts and price discrimination in competitive markets. *Journal of Political Economy*, 106(2), 395-422.

[11] Escobari, D., & Jindapon, P. (2014). Price discrimination through refund contracts in airlines. *International Journal of Industrial Organization*, 34, 1-8.

[12] GAO. (2001). Aviation competition: Restricting airline ticketing rules unlikely to help consumers. *Report to Congressional Committees*, U.S. General Accounting Office, Washington, DC.

[13] Gerardi, K. S., & Shapiro, A. H. (2009). Does competition reduce price dispersion? New evidence from the airline industry. *Journal of Political Economy*, 117(1), 1-37.

[14] Giaume, S., & Guillou, S. (2004). Price discrimination and concentration in European airline markets. *Journal of Air Transport Management*, 10(5), 305-310.

[15] Lazarev, J. (2013). The welfare effects of intertemporal price discrimination: an empirical analysis of airline pricing in US monopoly markets. *New York University*.

[16] Liu D. (2015). A Model of Optimal Consumer Search and Price Discrimination in the Airline Industry. *JMP*.

[17] Mathworks. *https://www.mathworks.com/help/gads/patternsearch-climbs-mt-washington.html*. online; accessed 3-October-2018.

[18] Pindyck, R. S., & Rubinfeld, D. L. (2018). *Microeconomics, 9th ed*. Pearson.

[19] Puller, S. L., & Taylor, L. M. (2012). Price discrimination by day-of-week of purchase: Evidence from the US airline industry. *Journal of Economic Behavior & Organization*, 84(3): 801-812.

[20] Rao, V. R. (2009). *Handbook of pricing research in marketing*. Edward Elgar Publishing.

[21] Shy, O. (2001). *The economics of network industries*. Cambridge University Press.

[22] Stavins, J. (2001). Price discrimination in the airline market: The effect of market concentration. *Review of Economics and Statistics*, 83(1), 200-202.

[23] Surry, P. (2016). Hidden city ticket opportunities are more common than you think. *https://media.hopper.com/research/hidden-city-ticket-opportunities-common-think*.

[24] Varian, H. R. (1980). A model of sales. *The American Economic Review*, 70(4): 651-659.

[25] Wang, Z., & Ye, Y. (2015). Hidden-city ticketing: the cause and impact. *Transportation Science*, 50(1), 288-305.