Essays on Estimation of Microeconomic Models

by

Quanquan Liu

Bachelor, Peking University, 2011

Master, Peking University HSBC Business School, 2014

Master, The University of Hong Kong, 2014

Submitted to the Graduate Faculty of

the Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Quanquan Liu

It was defended on

December 9, 2020

and approved by

Arie Beresteanu, Advisor University of Pittsburgh

Osea Giuntella, University of Pittsburgh

Federico Zincenko, University of Nebraska-Lincoln

Akshaya Jha, Carnegie Mellon University

Essays on Estimation of Microeconomic Models

Quanquan Liu, PhD

University of Pittsburgh, 2020

This dissertation consists of three studies.

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. In this paper, I build a structural model, collect empirical data, apply global optimization algorithms, and conduct counterfactual analysis to shed light on policy implications. I find that hidden city opportunity occurs only when airlines are applying a hub-and-spoke network structure. 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, for some routes airlines have the incentive to switch from hub-and-spoke network to a fully-connected one when there are more passengers informed of hidden city ticketing. Firms always result in lower expected profits, consumers and the whole society are not necessarily better off.

Global optimization without access to gradient information is a central task to many econometric applications as the tool to obtain maximum likelihood estimators for very complicated likelihood functions. In this work, we study the problem of coordination between the multiple "threads" of estimating gradient descent in order to pause or terminate unpromising threads early. We test our proposed methodology on both synthetic data and real airline pricing data, and compare with competitive methods including the genetic algorithm and pattern search. The numerical results show the effectiveness and efficiency of our proposed approach.

In my third work, I exploit large changes in the H-1B visa program and examines the effect of changes in H-1B admission levels on the likelihood that US natives major in STEM fields. I find some evidence that H-1B population adversely affect natives' choices in STEM fields when they enter the college and graduate from it. Female, Male and White subgroups have been negatively affected, and the native Asian subgroup suffer from the most dramatic crowd-out effect. Given that the H-1B population share had been more than doubled during

1992 to 2017, the probability of native Asian graduates majoring in STEM fields would be 2.56 percentage points larger.

Table of Contents

List of Tables

List of Figures

Preface

First and foremost, I would like to thank my advisor Arie Beresteanu, who has played an irreplaceable role in shaping my research interests, ability and methodology. Under his guidance and encouragement, I have learnt a lot and developed a passion for the field of applied microeconomics. He encourages me to pursue research questions that truly interest me, and is always there when I need helps or additional guidance in identifying and approaching some challenging topics. His support, guidance, wisdom, and encouragement all helped shape this thesis and much of my work, and they were all instrumental to my success in my PhD study.

I am also thankful to many other wonderful faculty members at Department of Economics, including Osea Giuntella and Federico Zincenko, who served on my thesis committee. I am grateful to the wonderful instructors of the PhD courses I took: Luca Rigotti, Marla Ripoll, Tymofiy Mylovanov, and Randall Walsh. I thank Akshaya Jha from Carnegie Mellon University, who served as my outside committee member, and provided valuable comments on my thesis. As a female student, I am truly fortunate to have Marla Ripoll and Stefania Albanesi as our graduates advisor and provided huge support and help during my PhD study.

Finally, I am eternally thankful my parents and my husband for their enduring and unconditional love and support. You are truly the best I have ever had in my life.

1.0 Paying More for a Shorter Flight? - Hidden City Ticketing

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 profits, while consumers and the whole society are not necessarily better off.

1.1 Introduction

Hidden city ticketing is an interesting pricing phenomenon occurring after the deregulation of the airline industry in 1978 ([84]). 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".

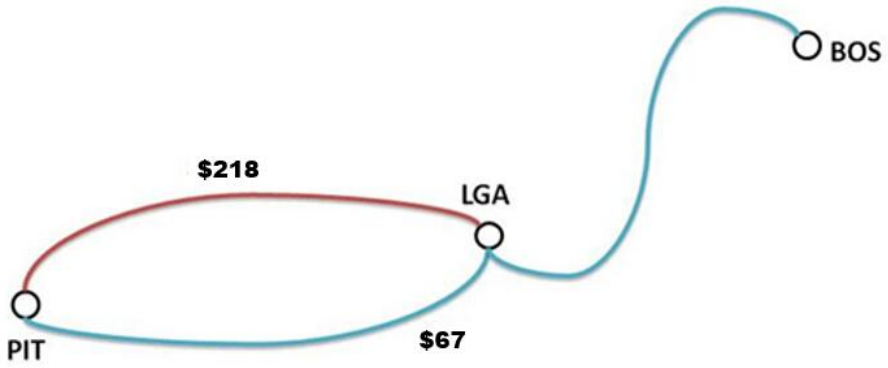


Figure 1: An example of hidden city ticketing.

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 the same first segment: they are operated by the same airline company, they departure at the same date, same airports and same time. However, the price of the indirect flight 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. It is "hidden" because literally, if we use Google Flights, Orbitz, Priceline, Kayak, or any other "normal" travel search tools to look for a flight from Pittsburgh to NYC, we will not be able to see the indirect flight above showing up in our search results.

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."

Similarly, American Airlines also claim that conducting 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". 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".

Meanwhile, members of Congress have proposed several bills, 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", trying to prohibit airlines from penalizing passengers for conducting hidden city ticketing ([38]). Furthermore, 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 have not suffered drastic losses due to this".

Therefore, whether hidden city ticketing should be legally prohibited or not, and what policy does the best for consumers, airline companies, and social welfare, still remain to be open questions. The fact is, 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 Chicago judge threw out United's lawsuit using the excuse that the founder "did not live or do business in that city". In contrast to the willingness of United, this lawsuit brought the search of key words "hidden city ticketing" to a peak (Figure 2).

Corresponding to this higher demand, nowadays there are more travel search tools specifically designed to achieve this task (Skiplagged, Tripdelta, Fly Shortcut, AirFareIQ, ITA

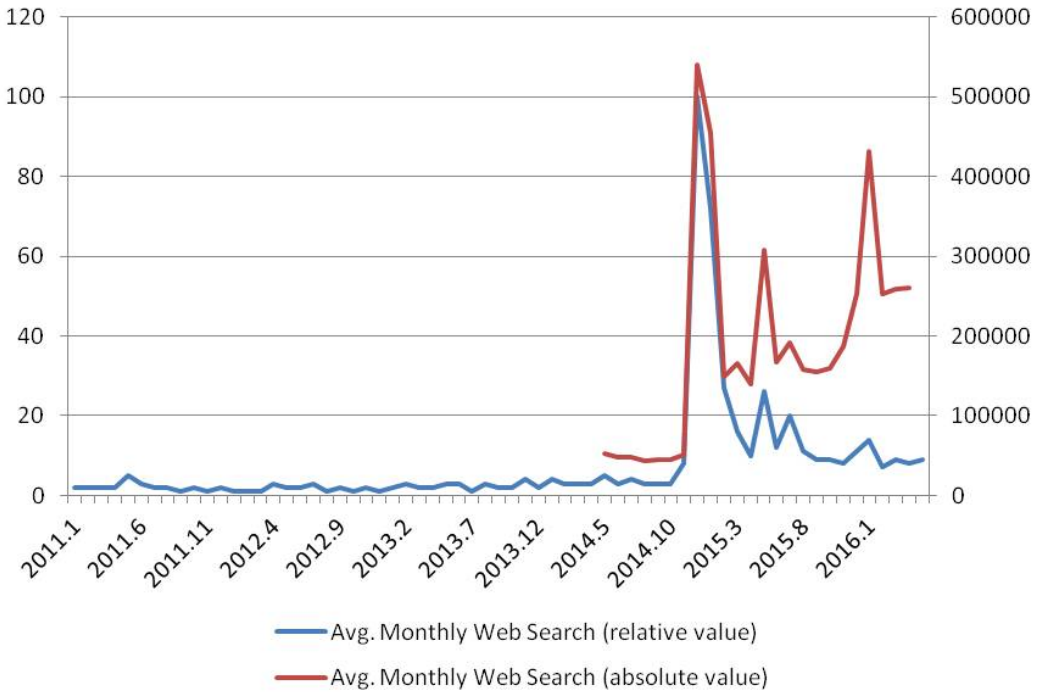


Figure 2: Average monthly web search data of hidden city ticketing. Data source for the relative value is Google Trends. Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. Data source for the absolute value is Google AdWords, unit is number of times.

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 [74], and derive several propositions based on that. Then I collect daily flights information by scraping the Skiplagged website 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.

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 does 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: Related works are summarized in Section 3.2. The structural model is introduced in Section 1.3, together with the propositions of short run impacts derived from it. Section 1.4 describes the data in details. Section 1.5 shows the estimation strategy and the MLE results. In Section 1.6 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 1.7 and Section 1.8 concludes.

1.2 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 Accountability Office ([38]) 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 ([79]) 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 [84], 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 has been a large 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, [12] has estimated a price elasticity of demand for tourists as -6.55, while that for business travelers is only -0.63. [71] 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). And [39] also confirm that the demand for business travelers is less price elastic than that of tourists, and through applying certain ticket restrictions, airline companies are able to distinguish between these two types. 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 ([34]), ticket restrictions such as Saturday-night stayover

requirements ([78]; [41]), refundable and non-refundable tickets ([36]), intertemporal price discrimination ([54, 52]), and even the day-of-the-week that a ticket is purchased ([67]). 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 [74] 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 [83] where the author shows some "detrimental externalities" that uninformed consumers suffer from due to the behavior of informed consumers.

1.3 The Model

Following [74], I assume that airlines are choosing from two different network structures: fully-connected network or hub-and-spoke network (Figure 3). Under 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 will apply an one-way traveling pattern instead of the two-way traveling pattern showed in 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 ([74]). Responding to the increased competition and to reduce flying costs, airlines started to cut the number of direct flights and reroute the passengers through a hub city. While in recent years, especially since late 1990s, with the expansion of low-cost carriers (LCCs), passengers started to show a higher aversion toward connecting flights, and fully-connected structure becomes more popular again ([12]).

Assume that there is only one airline serving the three cities, thus the firm charges monopoly airfares. Aircrafts are further assumed to have an unlimited capacity, thus there is only one flight on each route. C_2 denote the airline's cost per mile on any route j. This simplified cost pattern is referred as ACM cost (AirCraft Movement cost) in [74], and it is

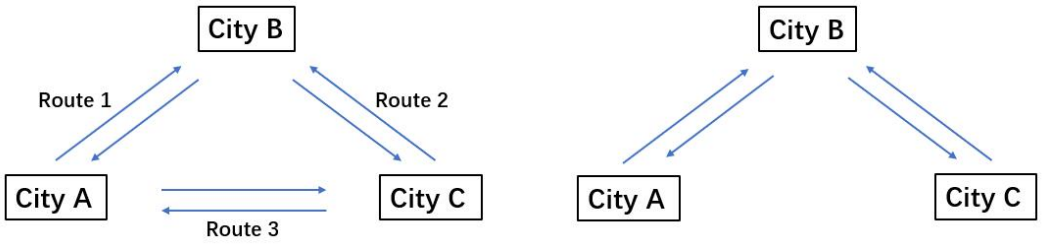


Figure 3: Left: fully-connected (FC) network. Right: hub-and-spoke (HS) network.

widely used in airline related literature. Cost pattern can be simplified because in airline industry, large percentage of costs are fixed before flights taking off, such as capital costs (renting gates for departure and arrival, landing fees), labor costs (hiring local staff), etc. ([38]) The costs of fuel account for approximately 15% of the total operation costs ([12]), while the marginal cost of airline seats is nearly negligible ([68]).

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

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

from consuming a good of quality q. Under the assumption of free disposal, he/she will get 0 utility if chooses not to fly. Utility decreases when price increases. And if a passenger values time more (i.e., with a larger λ), he/she will acquire a larger utility increase when switching from an indirect flight (with quality q_l) to a direct flight (with quality q_h). Furthermore, 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 opportunities is δ, and the fraction of uninformed passengers is $1 - \delta$. When hidden city opportunities exist (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. p_j denote the airfare on route j, and d_j denote the distance of route j.

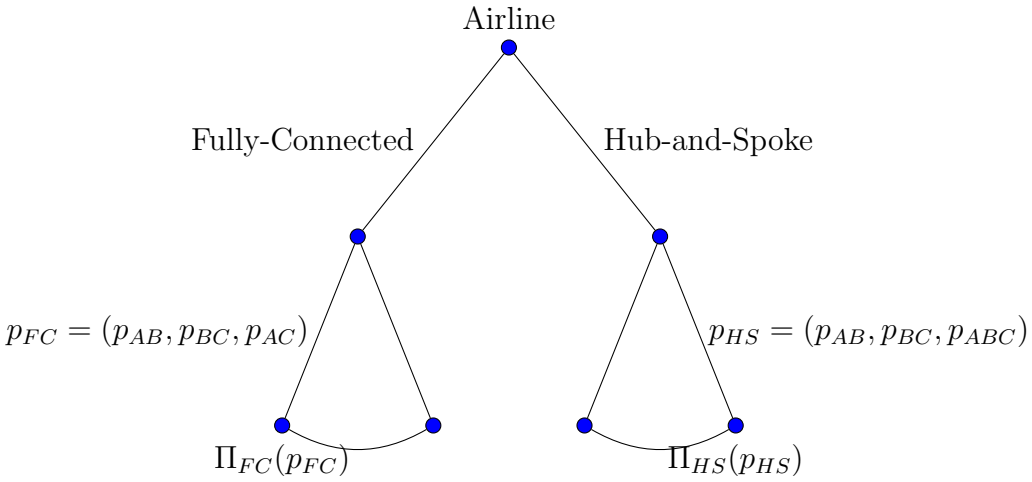


Figure 4: Airline chooses network structures and prices to maximize expected profits.

An airline chooses both network structures (fully-connected or hub-and-spoke) and prices $(p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ to maximize expected profits, as shown in Figure 4.

I will show later in this section that there exists an optimal choice set for the airline, and the choice set is unique. 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 choosing to fly is equal to:

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

1.3.1 Fully-Connected Network

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

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

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

Proposition 1. *Hidden city opportunity does not exist under fully-connected network structure.*

Consumer surplus is the difference between our willingness to pay and the price we actually being charged, which equals:

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

Adding them together, our total surplus under fully-connected network is:

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

No transaction fee is assumed under the setting, thus the prices we pay are equal to the prices airline receives, and both cancel out.

1.3.2 Hub-and-Spoke Network (without Hidden City Ticketing)

Given hub-and-spoke network structure, first consider the simple case when hidden city opportunities do not exist (i.e., $p_{AB} \leq p_{ABC}$). Under this circumstances, airline's expected profits are 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 2. *If the cost associated with maintaining route AC is sufficiently large, then the hub-and-spoke network is more profitable to operate than the fully-connected network for the monopoly airline.*

Proof. Proof of Proposition 2. 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 ($C_2 \cdot d_{AC}$, refers to the cost associated with maintaining route AC) is sufficiently large, hub-and-spoke network is more profitable. □

This is in accordance with the findings of previous literatures. [30], [23], [24], and [11] all confirm the cost economies of hubbing. Under a different framework, [74] also find that hub-and-spoke network is cost-saving if the fixed cost is large enough.

Similarly, consumer surplus equals:

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

Adding them together, total surplus is equal to:

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

1.3.3 Hub-and-Spoke Network (with Hidden City Ticketing)

Now consider the scenario when hidden city opportunities exist (i.e., $p_{AB} > p_{ABC}$). Firstly, is there a possibility that $p_{AB} > p_{ABC}$, in other words, are we paying more for a shorter flight sometimes? The answer is yes. To see why this might occur, recall that

$$p_{AB} = arg\max_{p}\ p \cdot \left[1 - \Phi_{\theta_{AB},\sigma_1^2}\left(\ln(\frac{p}{C_1 q_h d_{AB}})\right)\right],$$

$$p_{ABC} = arg\max_{p}\ p \cdot \left[1 - \Phi_{\theta_{ABC},\sigma_1^2}\left(\ln(\frac{p}{C_1 q_l d_{ABC}})\right)\right],$$

where $q_h > q_l$ and $d_{AB} < d_{ABC}$. For simplification, let $q_h d_{AB} = q_l d_{ABC}$ and $\sigma_1 = 1$, rewrite the problem as

$$\begin{aligned} p &= arg\max_{p}\ p \cdot \left[1 - \Phi_{\theta}\left(\ln(\frac{p}{C})\right)\right] \\ &= arg\max_{p}\ p \cdot \left[1 - \Phi\left(\ln(\frac{p}{C}) - \theta\right)\right], \end{aligned}$$

where C is a constant.

Let $f(p,\theta) = p \cdot \left[1 - \Phi\left(\ln(\frac{p}{C}) - \theta\right)\right]$, to find out the maximizer p^*, take derivative of $f(p,\theta)$ with respect to p and make it equal 0:

$$\begin{aligned} f_p(p,\theta) &= 1 - \Phi\left(\ln(\frac{p}{C}) - \theta\right) - \phi\left(\ln(\frac{p}{C}) - \theta\right) \\ &= 0. \end{aligned}$$

Let $g(p,\theta) = 1 - \Phi\left(\ln(\frac{p}{C}) - \theta\right) - \phi\left(\ln(\frac{p}{C}) - \theta\right)$, and take derivative of $g(p,\theta)$ with respect to θ, we get

$$\begin{aligned} g_\theta(p,\theta) &= \phi\left(\ln(\frac{p}{C}) - \theta\right) + \phi_\theta\left(\ln(\frac{p}{C}) - \theta\right) \\ &= \phi\left(\ln(\frac{p}{C}) - \theta\right)\left(1 - \ln(\frac{p}{C}) + \theta\right). \end{aligned}$$

Since $\phi\left(\ln(\frac{p}{C}) - \theta\right) > 0$, if $\theta > \ln(\frac{p}{C}) - 1$, we would have $g_\theta(p,\theta) > 0$, hence as long as $\theta_{AB} > \theta_{ABC}$, the optimal prices would be $p_{AB} > p_{ABC}$. Therefore, under some conditions, there is a possibility that $p_{AB} > p_{ABC}$, in other words, we are paying more for a shorter flight sometimes and hidden city opportunities exist. The underlying explanation is that airlines are pricing based on demand, rather than costs.

Comparing to Section 3.2 where there is no hidden city ticketing, the difference lies in the informed passengers who wish to fly directly from A to B. Under this circumstances, airline's expected profits are 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 3. *When airlines do not alter their choices of prices and network structures, hidden city ticketing does not necessarily decrease airline's expected profits.*

Proof. Proof of Proposition 3. Compare airline's expected profits with and without hidden city ticketing respectively, and compute the difference:

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

Note that $p_{ABC} < p_{AB}$ while $[1 - \Phi_{\theta_{AB},\sigma_1^2}(\ln(\frac{p_{ABC}}{C_1 \cdot q_h d_{AB}}))] > [1 - \Phi_{\theta_{AB},\sigma_1^2}(\ln(\frac{p_{AB}}{C_1 \cdot q_h d_{AB}}))]$. We find that although airline suffers from a loss when informed passengers are paying a lower price, it also obtains some gain when this lower price attracts more consumers to take the flight. How hidden city ticketing will affect airline's expected profits actually depends on the relative dominance of these two inequalities, and this conduct does not necessarily decrease airline's expected revenue. □

Consumer surplus equals:

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

Proposition 4. *When airlines do not alter their choices of prices and network structures, consumers are always better off when hidden city ticketing is allowed.*

Proof. Proof of Proposition 4. Compute the difference of consumer surplus with and without hidden city ticketing, we have

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

□

The increase in consumer surplus is composed of two different parts. Firstly, the existing informed passengers are now paying a lower price, which provides them extra utility gain. Secondly, some travelers who will not fly with the original price p_{AB} are now participating in this market activity, because they are informed of the lower price p_{ABC}. These new passengers also obtain utility gain, increasing the total consumer surplus.

Adding the producer surplus and consumer surplus together, we have the total surplus under the scenario of hub-and-spoke network structure with hidden city ticketing being 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 5. *When airlines do not alter their choices of prices and network structures, total social welfare always increase when hidden city ticketing is allowed.*

Proof. Proof of Proposition 5. Compute the difference of total surplus with and without hidden city ticketing, we have

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

□

Total surplus increase because compared to the original price p_{AB}, there are more travelers choosing to take the flight with the lower price p_{ABC}. Extra passengers obtain extra utility gain. Since I have assumed unlimited capacity for the aircrafts, the whole society benefit from this change.

With a full analysis of airline's expected profits under fully-connected network and hub-and-spoke network, with and without hidden city ticketing, we can now show that there

exists an optimal choice set for the airline to maximize its producer surplus, and the solution is unique.

Proposition 6. *Under the assumptions listed at the beginning of this section, there exists an optimal choice set* $(network, p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ *for the airline to maximize its expected profits, and the solution is unique.*

Proof. Proof of Proposition 6. According to our 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 $p_{AB} > p_{ABC}$.

Note that solving for the optimal choice set $(network, p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ is equivalent to firstly solving for the optimal price bundles $(p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ under fully-connected network and hub-and-spoke network respectively, and further compare $\Pi_{FC}(p_{AB}, p_{BC}, p_{AC})$

and $\Pi_{HS}(p_{AB}, p_{BC}, p_{ABC})$ to determine which joint choices of network structure and prices are optimal.

Solving for the optimal price bundle $(p_{AB}, p_{BC}, p_{AC}, p_{ABC})$ to maximize expected profits Π_{FC} and Π_{HS}, when there is no hidden city ticketing, 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 and make it equal 0:

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

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.

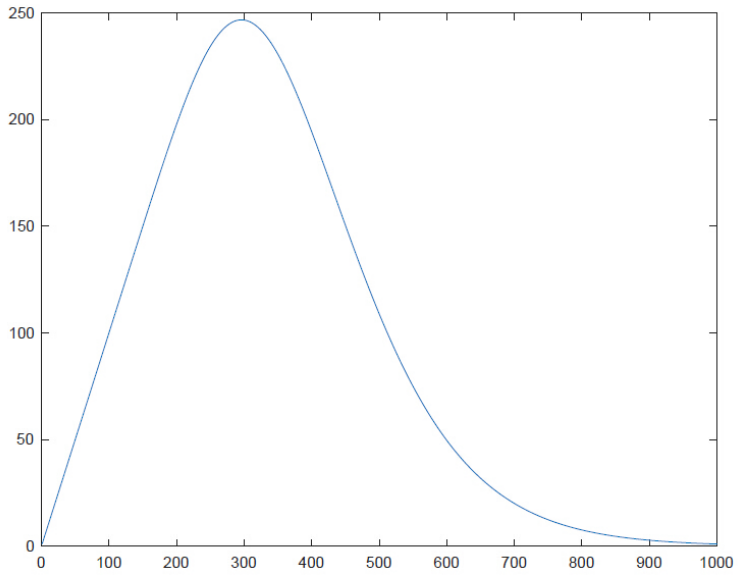


Figure 5: Illustration of function $f(p)$ with parameters
$\theta = 0.5,\ \sigma = 0.3,\ C = 10,\ q = 0.8,\ d = 30.$

Therefore, the optimal price bundle is solvable. 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 unimodal, I depict function $f(p)$ with parameters $\theta = 0.5,\ \sigma = 0.3,\ C = 10,\ q = 0.8,\ d = 30$, as shown in Figure 5 below.

When there is 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 and make it equal 0:

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

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. □

Note that Propositions 3, 4 and 5 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. This might be valid in the short run, while during a longer period, airlines should be able to realize the conduct of hidden city ticketing and adjust their optimal choices of prices and networks in response to the behavior. In such a scenario, obtaining a closed-form solution is challenging, and to reveal what would be the airline's optimal choice set with a changing proportion of informed passengers (changing δ) is even more difficult. Therefore, in the rest of the paper I will use numerical approach instead to solve for the optimal choices of airline with changing δ, and estimate the possible impacts of hidden city ticketing on welfare outcomes in the counterfactual analysis below.

1.4 Data

I have 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 have restricted my sample to the 133 busy commercial service airports identified by FAA. Only focusing on those 133 airports is reasonable because those airports

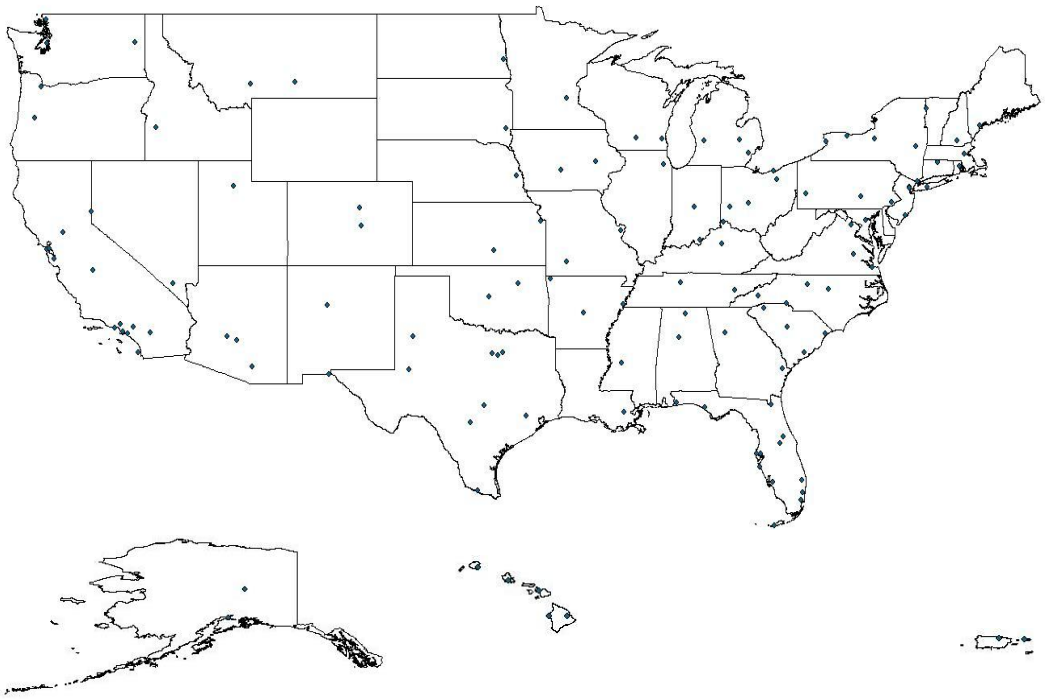


Figure 6: Distribution of busy commercial service airports around United States.

actually accounted for 96.34% of total passenger enplanements in 2014. The distribution of the busy commercial service airports around United States is shown in Figure 6. From the graph 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 has 16,142 routes (airport A to airport B) and 2,822,086 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 show some hidden city opportunities lying in the itineraries they operate.

To the best of my knowledge, there is no official definition of hidden city opportunity in existing literatures. [38] define that "a hidden-city ticketing opportunity exists 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 and listed 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 example being illustrated at the beginning of this paper belongs to the second scenario. And the second definition is also the one being defined in [84].

According to the daily flights data I have collected, these two definitions show similar magnitude with respect to hidden city opportunities. For example, among all the itineraries, the first definition indicates a total of 366,754 (13.00%) flights and 1,095 (6.78%) routes that exhibit possible hidden city opportunities. Those amounts of Definition 2 are 394,544 (13.98%) flights and 1,316 (8.15%) routes respectively. This magnitude is slightly smaller compared to the findings in [38], in which the authors find that among the selected markets for six major U.S. passenger airlines in their data, 17% provided such opportunities. Table 1

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%

shows the top 10 origin-destination pairs with most hidden city opportunities, which are the same under both definitions.

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 statistics help reveal the fact that hidden city ticketing might no longer be negligible nowadays and related research becomes necessary and valuable.

Recall that my primary data contains flights operated by 45 different airline companies, among which 11 companies shown 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 [79], 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.

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 is a typical operator of fully-

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%

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%)

connected network, this finding in the real data is in accordance with my previous proposition that hidden city opportunity does not exist under fully-connected network structure.

1.5 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 have applied global optimization algorithms, more specifically, Pattern Search to solve the MLE and estimate the parameters.

1.5.1 Sample Construction

To build my own sample, the first step is to retrieve ordered triplets (A-B-C) from the 133 busy commercial service airports of my primary dataset. The triplet needs to satisfy the following three conditions: 1) it must include direct flight from A to B; 2) it must include direct flight from B to C; 3) it must include either direct or one-stop indirect flight from A to C using B as the connection node.

In total, I have obtained 114,635 ordered triplets from my dataset that satisfy the conditions listed above. 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 computed based on 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, [17] and [20] built a tourism index at the MSA level based on the ratio of hotel income to total personal income. [23] and [78] assumed that the difference in January temperature between origin and destination cities could serve as a proxy for tourism. [39] 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 have constructed my own index based on [19] and data provided by TripAdvisor. [19] 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 [67] to distinguish between "leisure" and "mixed" routes. The shortage for this index is that it only includes data for each state and metropolitan statistical area, while city level data might be a better fit corresponding to the location of an airport. To solve this problem, I have also collected data from TripAdvisor (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 [19] should 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 [19] (both state-level and MSA-level) and get μ.

From Figure 7 we can see that the largest $\mu = 0.7450$ belongs to Dallas Fort Worth International Airport (DFW) in Texas, while Ellison Onizuka Kona International Airport (KOA) on the Island of Hawaii has the smallest $\mu = 0.0773$. In general, places that are more popular among tourists, such as Orlando, Puerto Rico and Hawaii, get the smaller μ(s). While places such as Dallas, Austin and Chicago that are more attractive to business

travelers have larger μ(s).

1.5.2 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 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

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

1.5.3 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 solve this MLE problem. 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 have 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 my log likelihood function. 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 does the Pattern Search algorithm operate? Pattern search applies polling method ([60]) 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.

1.5.4 Estimation Results

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

According to the estimation results, the informed passengers account for around 3.73% of the whole population. This proportion appeals to be trivial at first glance, but it is not surprising because those are the travelers who are not only informed of hidden city ticketing,

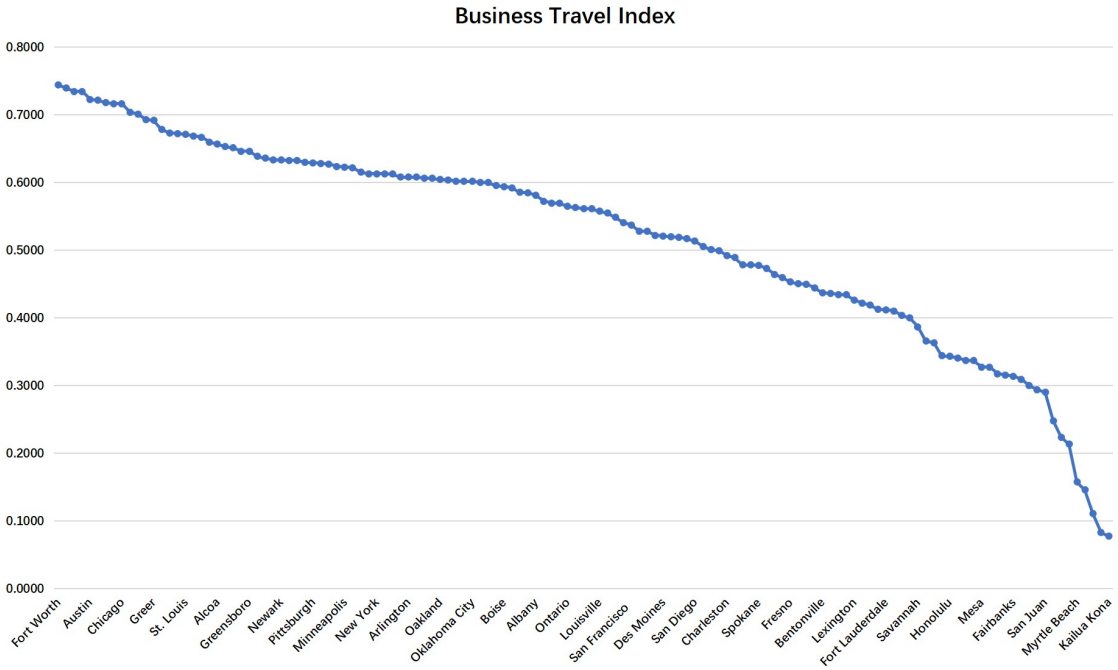


Figure 7: Business travel index for each airport as the destination city.

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

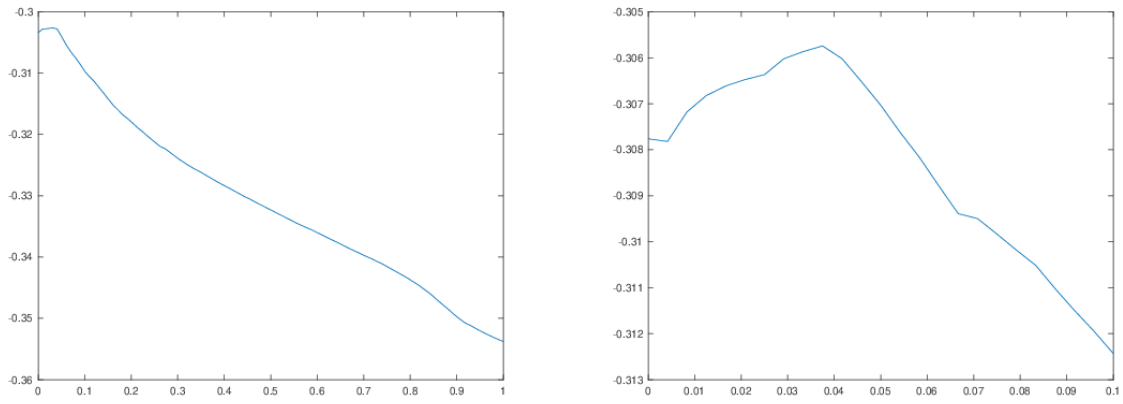


Figure 8: Up: Plot of log likelihood when δ varies from 0 to 1. Below: Plot of log likelihood when δ varies from 0 to 0.1 (zoom in).

but also exploiting those opportunities, and whose behavior in fact result in affecting the choices made by airlines. And in the counterfactual analysis section below, I will further show that even a small fraction of informed passengers will affect airline's choices of network structures and prices significantly.

The confidence interval for δ was constructed using the bootstrap method. I run the MLE for 1,000 times, and for each run, sample the entire data with replacement and construct a data set of equal size. The sample mean of the 1,000 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 8, I have 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.

Recall that in my theoretical model, I have assumed that there is only one airline serving the three cities, thus the firm charges monopoly airfares. [54] also made similar monopoly assumption in the paper, and corresponding to this assumption, the author refined the data and only paid attention to the routes with a single carrier operating one or two flights per day. Following this idea, I also define route AB, BC, AC, or ABC as monopoly if there

is only one single carrier providing services on that route. Refining my sample of ordered triplets according to this condition results in a subsample size of 36,645, comparing to the total sample size of 114,635 before. Applying the same estimation algorithm, I have solved the MLE problem again based on the monopoly subsample, and get an estimation of δ being equal to 0.0303. The result is not significantly different from the 0.0373 we obtain above from the whole sample.

1.6 Counterfactual Analysis

Based on our previous analysis, given a longer horizon, airlines should be able to adjust their prices and networks in response to hidden city ticketing, in order to maximize their expected profits. To reveal what would be the airline's optimal joint choices of prices and network structures when δ changes, and further estimate the possible impacts of hidden city ticketing on welfare outcomes, I have conducted several counterfactual analysis using numerical approach below. Will airline companies always suffer from revenue loss with hidden city ticketing? Will hidden city ticketing always benifit consumers and social welfare? Should government enact regulations to clearly prohibit or permit this booking ploy? My counterfactual experiments will help shed some light on those important policy implications.

Assume that the proportion of informed passengers (δ) increases from 0 to 100%, and airline companies always choose optimal prices under different network structures to maximize their expected profits when δ changes. After obtaining the optimal price bundle p^* under different networks, I compute the surplus of producer, consumer, and society according to our previous analysis in Section 1.3. Then I 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.

1.6.1 Fully-Connected Network Outperforms Hub-and-Spoke Network

Findings 1 *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 would be the ordered triplets MIA→SEA→COS (Miami International Airport to Seattle-Tacoma International Airport to Colorado Springs Airport). Figure 9 shows the surplus of producer, consumer, and society with different δ(s). The dotted lines are always horizonal because according to my model, hidden city ticketing will not affect the welfare outcomes under fully-connected network structure. It is clear that in this example, the dotted blue line is always above the solid blue one, regardless of the value of δ, which means that for airlines operating from Miami International Airport to Colorado Springs Airport, a direct flight always outperforms an indirect one through Seattle-Tacoma International Airport. This is not surprising because flying from Miami to Colorado through Seattle is counter intuitive.

When we plot the surplus in Figure 9, there is a pattern of kink, which is not uncommon and also found in other examples. Digging deep I find what happens at the kink is that airlines keep raising the price p_{ABC} in response to the increasing proportion of informed passengers, δ. Consumers benefit at first because more and more informed travelers are able to exploit the hidden city opportunities, pay lower prices and obtain extra utility. However, when the kink point is reached, p_{ABC} hits the magnitude of p_{AB} and hidden city opportunities disappear. The informed passengers can no longer obtain extra utility, while since the new p_{ABC} turns out to be higher than the original price without hidden city ticketing, those passengers flying from A to C through B also get hurt. This is similar to what is called "detrimental externalities" in [83], in which the author also found that sometimes more informed consumers would cause the price paid by uninformed consumers to increase. This finding also helps confirm the concern mentioned in [38] that allowing hidden city ticketing might lead to unintended consequences, including higher prices.

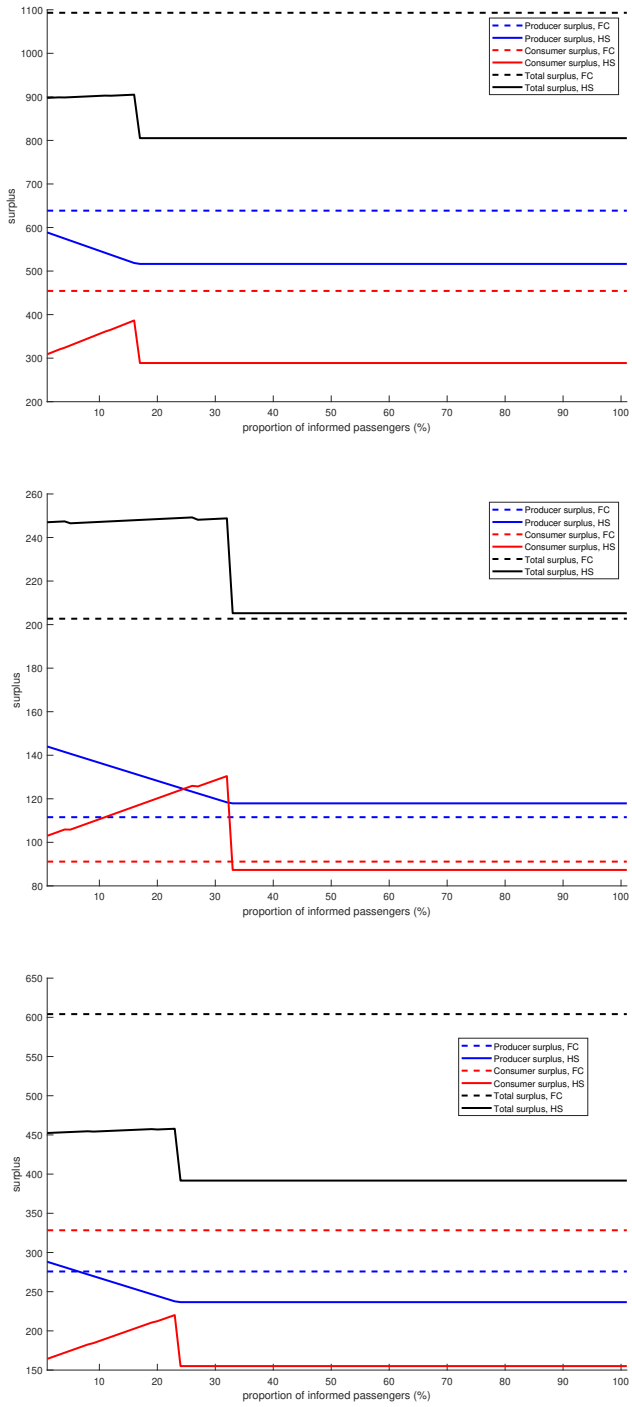


Figure 9: Surplus for MIA to SEA to COS (top left), CID to DTW to MSN (top right), AUS to JFK to RDU (bottom) when δ changes.

1.6.2 Hub-and-Spoke Network Outperforms Fully-Connected Network

Findings 2 *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.

Contradict to the previous finding, sometimes the hub-and-spoke network structure always does a better job achieving higher revenue compared to fully-connected network. 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 9 shows the surplus of producer, consumer, and society in this case when δ varies.

We can see that the solid blue line is always above the dotted blue one, regardless of the value of δ, which means that for airlines flying from The Eastern Iowa Airport to Dane County Regional Airport in Madison, an indirect flight through Detroit Metropolitan Airport always outperforms a direct flight. This 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.

1.6.3 Switch from Hub-and-Spoke Network to Fully-Connected Network

Findings 3 *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 *gets large.*

A more interesting story lies in the cases remained: hub-and-spoke network structure is more profitable when δ is small, but becomes gradually outperformed by fully-connected network when there are more and more informed passengers. In other words, for some specific routes, airlines have the incentive to switch from one network structure to another, and δ will affect companies' network choices. This finding could be supported by what we called "dehubbing" phenomenon in recent years ([11]). For example, Delta closed its Dallas-Fort Worth International Airport (DFW) hub in year 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 shows the surplus of producer, consumer, and society in this case when δ varies.

We can see that the solid blue line crosses the dotted blue one at the point when δ is around 6%, which means that when δ is smaller than the threshhold, airline would pursue hub-and-spoke network structure. While when there are more and more informed passengers and δ crosses the threshhold, airline has the incentive to switch from the hub-and-spoke network to fully-connected network, and this decision will also affect both consumer surplus and total surplus dramatically. In this example, after the airline company making the change, both consumer surplus and total surplus increase a lot, which refer to the increase from the solid red, black lines to the dotted red, black lines respectively. But this is not always the case, and we will see more details later in this paper.

The crossing point varies for different ordered triplets. This is because different routes have different characteristics and attract different types of travelers. Some routes would be quite "sensitive" to hidden city ticketing and airlines operating on those routes would switch from hub-and-spoke network to fully-connected network when δ is relatively small. Some routes would have operating airlines changing their network choices only when the amount of informed passengers are large enough. And we have already known that sometimes airlines will never change to the fully-connected network (Findings 2), while in other cases they will stick to the fully-connected network from the very beginning (Findings 1). To have a more complete idea about the impact of different δ(s) on airlines' network choices, I can always depict the graph of surplus for every ordered triplet in my data. However, it is impossible to display all those figures here (recall that I have as many as 114,635 data points in total). Therefore, in Figure 10 I have plotted the distribution of all the crossing points when δ changes.

We can see that even a small δ matters. Airlines' choices can be affected significantly even with a quite small proportion of informed passengers. For example, airlines would switch from hub-and-spoke network to fully-connected network on nearly 1,000 routes with only 1% of informed passengers, and further change their choices on another 900 routes if

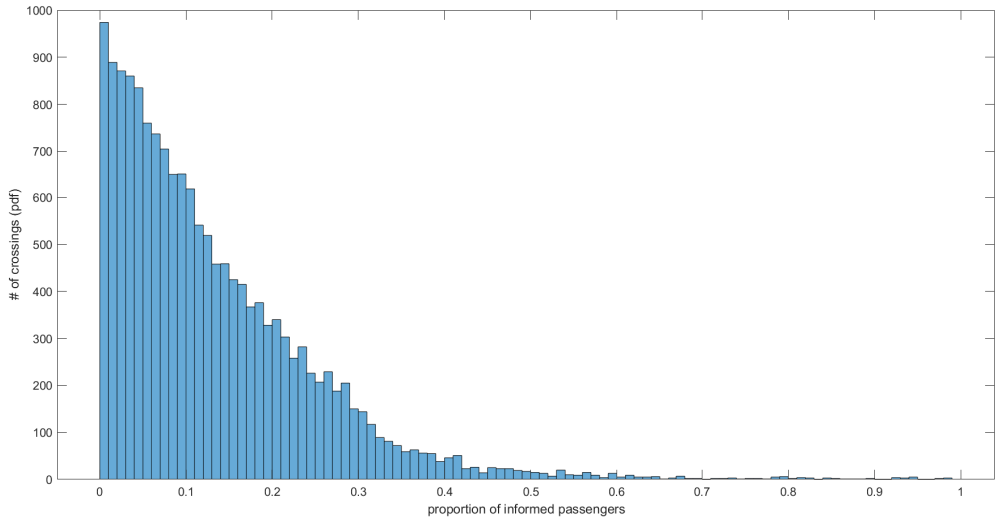


Figure 10: Distribution of crossings when δ changes (pmf).

the proportion increases to 2%. Recall that we have obtained an estimation of $\delta = 3.73\%$ in Section 1.5, which appeals to be trivial at first glance, but in fact, 3% of informed passengers could affect airlines' choices of network structures on approximately 2,700 routes (out of 16,089 in my whole sample), and this amount of routes being affected would increase to around 3,600 when the proportion of informed passengers increase to 4%. To make this illustration clearer, I have also drawn the cumulative distribution function of the crossing points when δ varies from 0 to 1, and obtain my next finding from the following Figure 11.

Findings 4 *Airlines have the incentive to switch from hub-and-spoke network to fully-connected network for half of the routes when there are approximately 10% of informed passengers, and for 75% of the routes when δ is only around 19%.*

Recall that after airlines changing their choices of network structures, consumers and the whole society are not always better off. After comparing the consumer surplus and total surplus before and after the change for all those 16,089 routes, I am able to further conclude that:

Findings 5 *If airlines switch from hub-and-spoke network to fully-connected network, un-*

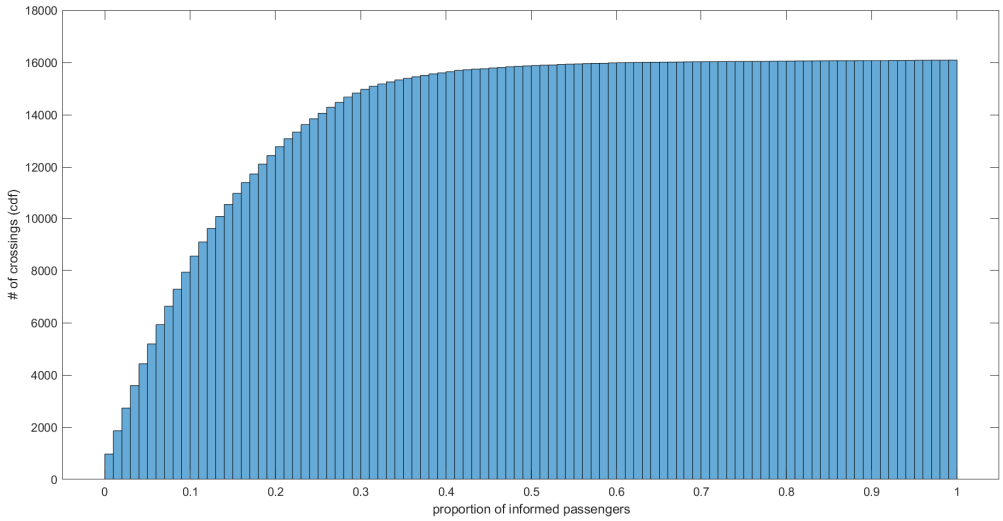


Figure 11: Distribution of crossings when δ changes (cdf).

der 11,458 cases (71.22%) consumer surplus is going to increase, and under 11,128 (69.17%) cases total surplus is going to increase.

In other words, unlike what could be derived from the theoretical model when airlines do not alter their choices of network structures and airfares in the short run, during a longer horizon, firms would actively react to the conduct of hidden city ticketing and change their optimal choices. This will result in different welfare outcomes for producer, consumer, and whole society compared to the propositions in Section 1.3. What I have found in my counterfactual analysis is that, during this process, firms always result in lower expected profits, while consumers and the whole society are not necessarily better off. One possible explanation would be that the conduct of hidden city ticketing weakens the airlines' ability of price discrimination. Therefore, enacting a simple regulation for prohibiting or permitting the conduct of hidden city ticketing would be difficult.

1.7 Discussion

There are some limitations and future questions remained in this paper. Firstly, I have made a critical assumption in my model that there is only one airline serving the three cities, thus the firm charges monopoly airfares. The monopoly assumption is not uncommon in airline related literatures, but some researchers believe that after the deregulation, the US airline industry should be characterized as being hightly oligopolistic ([74]). And in [38], the authors also find that "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". In other words, besides the factors I have 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 effect, 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 the prices for the origin-hub passengers. Since there are business travelers who favor the origin-hub route and appeal to be price-inelastic, this market power of raising prices could result in higher profits for the hubbing airlines. ([17], [18])

Another interesting question raised by [83] is that does it pay to be informed? If this is true, a better way to take this into consideration might be assuming that it is possible to become fully informed by paying a fixed cost C. Conducting hidden city tickeing is definitely costly. Passengers are "threatened" by the airline companies and need to "bear some risk" to conduct this behavior. As I have quoted from the contract of carriage in Section 3.1, consumers will be penalized if being caught. And hidden city ticketing might be treated as "unethical" and a breach of the contract between passengers and the airlines. There are also possible negative externalities such as causing delays of the other passengers because of waiting and double checking baggages. Furthermore, the condition of conducting hidden city ticketing is also highly restrictive. For example, if you have luggage that is not carry-on, you are not able to leave the flight earlier without picking up your bag. Normally, your checked baggage will be delivered to your final destination directly rather than to your connection

city. Also, 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. Besides, you might need to bear the risk that you are switched to another flight because of initial flight being cancelled or overbooked, with the same origin and destination airports, but bypass the connection city. Therefore, cost incurred to be informed might be a reasonable assumption when studying hidden city ticketing in the future, while measuring this cost would still be challenging.

Furthermore, airlines claim in the news that in reaction to the booking ploy of hidden city ticketing, they might choose to charge more on flights, stop offering some flights, and re-calibrate their no show algorithms. My analysis successfully predicts the increase in airfares of some flights. But since I have assumed that airline companies are choosing between fully-connected network and hub-and-spoke network, I do not provide an outside option for airlines to stop offering the flights for certain routes. Instead of raising the prices of flights to eliminate hidden city ticketing, another possibility is that the airlines could stop serving those defective routes. This is also another major concern in [38] that allowing hidden city ticketing might result in unintended decreasing service. Including this outside option could help make the analysis more complete, although the magnitude might be difficult to evaluate without a good measure of the costs. Another concern raised by purchasing hidden city tickets is related to logistics and public-safety. When hidden city ticketing becomes more popular, airlines might need to re-calibrate their no show algorithms. They might have the incentive to oversell more, which is an act that could turn problematic and expensive if the estimates are wrong. Unfortunately, I did not find the dataset of airlines' oversales. Related statistics provided by United States Department of Transportation are numbers of passengers boarded and denied boarding by the U.S. Air Carriers. Figure 12 shows the percentage of passengers being denied boarding by the U.S. Air Carriers because of oversales, both voluntarily and involuntarily, from year 1990 to 2019.

From Figure 12 we can see a declining trend of percentage of passengers being denied boarding because of oversales in recent years, which indicates that the concern of possible oversales raised by hidden city ticketing might be subtle.

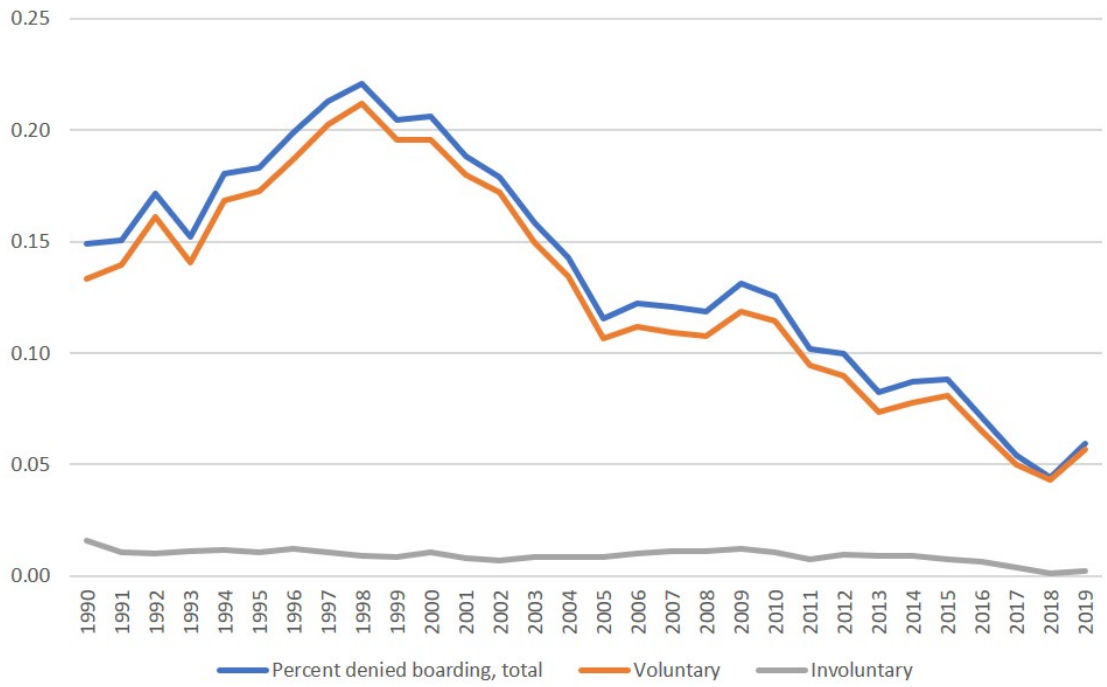


Figure 12: Percentage of passengers denied boarding by the U.S. air carriers, 1990 to 2019.

1.8 Conclusion

To conclude, this paper aims at analyzing the possible cause and impact of hidden city ticketing. To achieve this goal, I have constructed a structural model, collected innovative data, applied global optimization algorithm to solve the MLE, and conducted counterfactual analysis. I find that hidden city ticketing occurs only when airline companies are applying a hub-and-spoke network structure. And airlines apply hub-and-spoke network rather than fully-connected network in order to reduce their operation costs. When airlines are not aware of hidden city ticketing, hence do not alter their choices of prices and network structures in the short run, we can derive from the theoretical model that, 1) hidden city ticketing does not necessarily decrease airline's expected profits, since the lower price also attracts more passengers to take the flight; 2) consumers are always better off when hidden city ticketing is allowed; 3) total social welfare always increase when hidden city ticketing is allowed.

During a longer horizon, firms would actively react to the conduct of hidden city ticketing and freely change their optimal choices of network structures and airfares. Under this circumstances, based on the counterfactual analysis I have conducted, I find that 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 cases, airlines' expected profits are larger under hub-and-spoke network when there are less informed passengers, while fully-connected network becomes more profitable when more and more passengers starting to exploit 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 approximately 10% of informed passengers, and for 75% of the routes when informed passengers increase to around 19%; 4) if airlines change their network choices because of hidden city ticketing, firms are suffering from revenue loss, while consumers are not always better off (28.78% of the cases consumer surplus will decrease), and total social welfare is not always larger neither (30.83% of the cases total surplus will decrease). Therefore, enacting a simple regulation for prohibiting or permitting the conduct of hidden city ticketing would be difficult, because welfare outcome varies route by route.

2.0 Maximum Likelihood Optimization via Parallel Estimating Gradient Ascent

Global optimization without access to gradient information is a central task to many econometric applications as the tool to obtain maximum likelihood estimators for very complicated likelihood functions. The estimating gradient descent framework is particularly popular, which uses local functional evaluation to build gradient estimates and perform gradient descent from multiple initial points. In this work, we study the problem of coordination between the multiple "threads" of estimating gradient descent in order to pause or terminate unpromising threads early. The high-level idea is to make predictions, either conservative or aggressive, on the potential progress of each estimating gradient descent threads and to compare them with the progress on other threads. We also test our proposed methodology on both synthetic data and real airline pricing data, and compare with competitive methods including the genetic algorithm and the pattern search algorithm. The numerical results show the effectiveness and efficiency of our proposed approach.

2.1 Introduction

Maximum-likelihood (ML) estimation is the workhorse for a wide range of inferring tasks of econometric modeling. Mathematically, given model $p(x;\theta)$ with unknown parameter of interest $\theta \in \Theta \subseteq \mathbb{R}^d$ and collected data sample $\{x_i\}_{i=1}^N$, the ML estimation problem can be formulated as

$$\max_{\theta\in\Theta} \frac{1}{N}\sum_{i=1}^{N} \log p(x_i;\theta). \tag{1}$$

In the rest of this paper, we also abbreviate $F_i(\theta) := \log p(x_i;\theta)$ and $F(\theta) := \frac{1}{N}\sum_{i=1}^N F_i(\theta)$. The optimization question of Eq. (1) is then equivalent to $\max_{\theta\in\Theta} F(\theta)$.

The primary focus of this paper is on the *computation* of approximate solutions of the ML estimates (1) under challenging scenarios when the underlying model $p(\cdot;\theta)$ is very

complicated and has several undesirable properties. Some important challenges include but are not limited to:

1. **Non-concavity and non-unimodality of log-likelihood**: the log-likelihood function $\log p(x_i; \cdot)$ might not be concave or even uni-modal with respect to the unknown parameter θ, making "local search" type methods such as mountain-climbing or gradient ascent difficult to find *global* optima of Eq. (1). Instead, it is very likely that these methods would stuck in local minima or saddle points;
2. **Inaccessible first-order information**: many optimization methods, such as gradient ascent or Newton's method, requires access to first-order or even second-order derivatives $\nabla_\theta \log p(x_i; \theta), \nabla^2_\theta \log p(x_i; \theta)$. Unfortunately, for many complicated econometric models, even if such derivatives exist, they still cannot be easily computed in closed forms. This prevents straightforward adoptions of famous continuous optimization algorithms to solve Eq. (1);
3. **Noisy zeroth-order evaluation**: in some scenarios, even the log-likelihood function $\log p(x_i; \theta)$ itself cannot be evaluated or computed without error, given data x_i and a hypothetical parameter θ. For example, in econometric models involving game theoretical process with unknown parameters, the computation of $\log p(x_i; \theta)$ requires multiple Monte-Carlo samples and cannot achieve arbitrary levels of accuracy;
4. **Curse of dimensionality for multiple parameters**: while nonparametric estimation of the log-likelihood function naturally leads to an approximate optimization algorithm for Eq. (1), such an approach suffers from the *curse of dimensionality* when there are more than one parameters (i.e., $d > 1$). Even when d is moderately large (e.g., $d \in [5, 10]$), estimating the entire log-likelihood function could already lead to unacceptable level of computations.

In this paper, we propose an algorithm framework which we call "parallel estimating gradient ascent". Our algorithm has the following properties:

(a) **Parallelism**: the algorithm runs *in parallel* several computing threads simultaneously, with vastly different initial points. This partially avoids the problem of being stuck in local optima/saddle points, as the different computing threads could well lead/converge

to different local minima;

(b) **Gradient estimation**: our method *estimates* the first-order derivative of the log-likelihood function only with access to noisy evaluation of $\log p(x_i; \theta)$, making the proposed method applicable to a wide range of econometric models;

(c) **Coordination of computing threads**: instead of running all parallel computing threads with the same pace, we design "selection rules" and "stopping rules" to carefully *coordinate* the different computing threads, so that promising computing threads are devoted with more computing resource and unpromising threads are terminated early without wasting more computing time.

The rest of this paper is organized as follows: In Sec. 2.2 we give accounts to related works. The description of our proposed algorithm, as well as its several components and convergence guarantees, are given in Sec. 2.3.

2.2 Related Works

The idea of using iterative methods to solve stochastic optimization questions without first-order information is a well-studied topic in the literature of mathematical optimization, machine learning and computational statistics. Estimating gradient descent/ascent originates from the works of [49, 16], which was later studied and explored in [62, 37, 13, 1, 73]. Our proposed approach also resembles "zeroth-order trust region" algorithms, studied in the works of [10, 15, 77, 66, 32, 31]. The majority of this line of work assumes the objective function to be optimized is *convex* (or concave, for maximization problems). An exception is the work of [40], which considered non-convex objectives and studied how fast iterative methods converge to *stationary points* of the said objectives.

Apart from iterative or estimating gradient type methods, many other heuristic algorithms are also well-known for such global optimization questions considered in this paper. Examples include the genetic algorithm [51, 85], simulated annealing [82], pattern search [80, 53], as well as modern approaches such as Bayesian optimization [76] and hierarchical optimistic optimization [25, 26].

The idea of running several computing threads coordinating them appropriately has also been explored in [2] in which bandit optimization problems are solved by running multiple candidate algorithms in parallel, and in [59] to solve multi-armed bandit problems subject to an unknown amount of adversarial corruption.

2.3 Algorithm Description

The pseudo-code description of the proposed algorithm framework is given in Algorithm 1. The algorithm consists of three major components: ThreadCoordination, GradientEstimation and ThreadStopping, which we describe in further details below and in subsequent sections:

- ThreadCoordination: this component aims at the selection (at iteration t) of an active thread $j \in \mathcal{A}_t$, where $\mathcal{A}_t$ is the subset of all active threads at time t (see also the description of ThreadStopping for the interpretation and construction of active subsets). While the uniform distribution $U(\mathcal{A}_t)$ is the most widely used (which spreads the computing resources evenly among all remaining active threads), other distributions could be developed that favor more "promising" threads. Further details and discussion are given in Sec. 2.3.1.
- GradientEstimation: this component aims at the (approximate) computation of the first-order derivative $\nabla_\theta F(\theta)$ using only noisy evaluations of $F(\theta + \delta)$, where $\delta \in \mathbb{R}^d$ is a small d-dimensional perturbation vector. The computation is enabled by first-order Taylor expansions of the objective function F centered at θ. Further details and discussion are given in Sec. 2.3.2.
- ThreadStopping: this component identifies unpromising threads which are not possible to lead to or converge to good solutions. Such unpromising threads are identified through comparison with the results from other threads, and are subsequently removed from consideration in future iterations. The subset $\mathcal{A}_t$ is defined as the set consisting of all active threads at iteration t, which forms the support of the thread distribution used

Algorithm 1 The meta-algorithm framework.

1: Let $\hat{\theta}_0^{(1)}, \hat{\theta}_0^{(2)}, \cdots, \hat{\theta}_0^{(J)}$ be the initial parameter estimates of J computing threads; initialize also $\mathcal{J} = \{1, 2, \cdots, J\}$ as the set of active threads;

2: **for** $\tau = 0, 1, 2, \cdots$ **do**

3: Select $j_\tau \in \mathcal{J}$ using the ThreadCoordination component;

4: Compute $g_\tau = \hat{\nabla}_\theta F(\hat{\theta}_\tau^{(j_\tau)})$ using the GradientEstimation component;

5: Perform *projected ascent* step for thread j_τ: $\hat{\theta}_{\tau+1}^{(j_\tau)} = \mathcal{P}_\Theta(\hat{\theta}_\tau^{(j_\tau)} - \eta_\tau \hat{\nabla}_\theta F(\hat{\theta}_\tau^{(j_\tau)}))$, where η_τ is a certain step size and $\mathcal{P}_G(\cdot) = \arg\min_{\theta \in \Theta} \| \cdot - \theta \|_2$;

6: Eliminate threads in $\mathcal{J}$ using the ThreadStopping component;

7: **end for**

in the ThreadCoordination component. Further details and discussion are given in Sec. 2.3.3.

2.3.1 The ThreadCoordination Component

The ThreadCoordination component, as suggested by its name, determines (potentially randomly) at each time stamp t the particular computing thread j to be pursued for the next iteration.

For clarity, we assume the algorithm is currently at time t (i.e., a total number of t observations have already been collected), and thread $j \in J$ is at parameter estimate $\hat{\theta}_t^{(j)}$. We also use $\mathcal{T}_{j,t}(\Delta_t)$ to denote the time stamps for the previous Δ_t times thread j is selected, up to time stamp t. We shall also assume that t is not too small. (When t is very small, meaning that the optimization algorithm just starts, there is very little information regarding the performance of each computing threads and hence the threads to pursue should be selected uniformly at random from all J threads.) We use the following three aspects to decide on which computing thread is to be pursued next:

1. *Current performance*: the *current* performance of each computing thread is of the utmost importance, since a thread that already performs well (i.e., attaining parameter estimates with good objective values) is likely to further push the performance of the

entire algorithm/system. For a particular thread j, its current performance can be estimated as the average objective values over the last Δ_t time stamps at which thread j is selected, or more specifically

$$\mathsf{CP}(j,t) := \frac{1}{\Delta_t} \sum_{\tau \in \mathcal{T}_{j,t}(\Delta_t)} \hat{F}(\hat{\theta}_\tau^{(j)});$$

2. *Estimated progress*: a thread that can progress fast (i.e., rapidly increasing the objective value of the parameter estimates) should be pursued more frequently since they will likely deliver fast progress for the overall algorithm/system performance as well. The estimated progress of thread j can be obtained by comparing the its performance for the previous Δ_t and $2\Delta_t$ time stamps during which thread j is selected, or more specifically

$$\mathsf{EP}(j,t) := \frac{1}{\Delta_t} \sum_{\tau \in \mathcal{T}_{j,t}(\Delta_t)} \hat{F}(\hat{\theta}_\tau^{(j)}) - \frac{1}{\Delta_t} \sum_{\tau \in \mathcal{T}_{j,t}(2\Delta_t) \setminus \mathcal{T}_{j,t}(\Delta_t)} \hat{F}(\hat{\theta}_\tau^{(j)});$$

3. *Volatility*: the "volatility" of a thread measures how stable/volatile of the quality of the parameter estimates obtained by the thread. It can be estimated by the sample standard deviation of the performance for the pervious Δ_t time stamps during which thread j is selected, or more specifically

$$\mathsf{V}(j,t) := \sqrt{\frac{1}{\Delta_t} \sum_{\tau \in \mathcal{T}_{j,t}(\Delta_t)} (\hat{F}(\hat{\theta}_\tau^{(j)}) - \mathsf{CP}(j,t))^2}.$$

Once the statistics $\mathsf{CP}(j,t), \mathsf{EP}(j,t), \mathsf{V}(j,t)$ are computed for each thread j, an aggregated statistic $\mathsf{A}(j,t)$ is calculated as

$$\mathsf{A}(j,t) = \kappa_{\mathsf{CP}} \mathsf{CP}(j,t) + \kappa_{\mathsf{EP}} \mathsf{EP}(j,t) + \kappa_{\mathsf{V}} \mathsf{V}(j,t),$$

where $\kappa_{\mathsf{CP}}, \kappa_{\mathsf{EP}}, \kappa_{\mathsf{V}} \geq 0$ are pre-determined weight parameters that carefully balance the three aspects we discussed above. The larger the value of $\mathsf{A}(j,t)$ is, the more promising the thread j is deemed by our algorithm at time t. The thread selection/coordination rule is then designed as

$$\Pr(\text{select thread } j \text{ at time } t) = \frac{\exp\{\mathsf{A}(j,t)\}}{\sum_{j' \leq J} \exp\{\mathsf{A}(j',t)\}},$$

which selects computing thread j at random with probability positively correlated with its score $\mathsf{A}(j,t)$.

Algorithm 2 The GradientEstimation component/procedure.

1: **Input**: solution point $\theta \in \mathbb{R}^d$, probing radius $\delta > 0$, number of probing points $m \in \mathbb{N}$;

2: **Output**: $\hat{\nabla}_\theta F(\theta) \in \mathbb{R}^d$, an estimate of $\nabla_\theta F(\theta)$.

3: Sample $u_1, \cdots, u_m$ uniformly at random from the unit sphere $\{u \in \mathbb{R}^d : \|u\|_2 = 1\}$;

4: For each probing vector u_i, $i \in \{0, 1, \cdots, m\}$, collect noisy function evaluation y_j, y'_j at θ and $\theta + \delta u_i$, respectively such that $\mathbb{E}[y_j|\theta, x_j] = F(\theta)$ and $\mathbb{E}[y'_j|\theta, x_j] = F(\theta + \delta u_j)$;

5: Find $\hat{\nabla}_\theta F(\theta)$ as the least-squares estimation

$$\hat{\nabla}_\theta F(\theta) = \arg\min_{g \in \mathbb{R}^d} \frac{1}{m} \sum_{j=1}^{m} \left| \frac{y'_j - y_j}{\delta} - \langle u_j, g \rangle \right|^2.$$

2.3.2 The GradientEstimation Component

The Gradient Estimation component aims at the approximate computation of the first-order derivative $\nabla_\theta F(\theta) = \frac{1}{N} \sum_{i=1}^{N} \nabla_\theta \log p(x_i; \theta)$. A pseudo-code description of GradientEstimation is given in Algorithm 2.

As motivated in the introduction, such first-order derivatives $\nabla_\theta F(\theta)$ cannot be directly computed in closed forms. Instead, given a hypothetical parameter θ' (possibly different from θ at which the derivative $\nabla_\theta F(\theta)$ is sought), one can compute a *noisy* evaluation of $F(\theta') = \frac{1}{N} \sum_{i=1}^{N} \log p(x_i; \theta')$. With many such noisy evaluations at different "probing" positions θ', a local linear model can be constructed via first-order Taylor expansions and least-squares estimators are employed to find an estimator of $\nabla_\theta F(\theta)$.

The following lemma gives an upper bound on the estimation error of $\hat{\nabla}_\theta F(\theta)$ under the assumption that the gradients of the log-likelihood to be estimated, $\nabla F(\cdot)$, is Lipschitz continuous.

Lemma 1. *Suppose $\nabla_\theta F(\cdot)$ is L-Lipschitz continuous, meaning that $\|\nabla_\theta F(\theta) - \nabla_\theta F(\theta')\|_2 \leq L\|\theta - \theta'\|_2$ for all θ, θ'. Suppose also that $\mathrm{Var}(y_j), \mathrm{Var}(y'_j) \leq \sigma^2$ for some $\sigma > 0$, and $m \geq 8d \ln(d/\delta)$ for some $\delta \in (0, 1/2)$. Then with probability at least $1 - \delta$, the estimation error $\hat{\nabla}_\theta F(\theta) - \nabla_\theta F(\theta)$ can be decomposed as*

$$\hat{\nabla}_\theta F(\theta) - \nabla_\theta F(\theta) = \beta + \zeta,$$

where $\|\beta\|_2 \leq L\delta$ *almost surely and* $\zeta \in \mathbb{R}^d$ *is a random vector satisfying* $\mathbb{E}[\zeta|\theta] = 0$ *and* $\mathbb{E}[\zeta^\top \zeta|\theta] \leq 4d^2/(m\delta^2)$.

Proof. Proof of Lemma 1. For notational simplicity denote $y_j = F(\theta) + \varepsilon_j$ and $y'_j = F(\theta + \delta u_j) + \varepsilon'_j$ where $\mathbb{E}[\varepsilon_j|\theta, u_j] = \mathbb{E}[\varepsilon'_j|\theta, u_j] = 0$ and $\mathrm{Var}[\varepsilon_j|\theta, u_j], \mathrm{Var}[\varepsilon'_j|\theta, u_j] \leq \sigma^2$. By the mean-value theorem, there exists $\tilde{u}_j = \lambda\delta u_j$ for some $\lambda \in (0, 1)$ such that

$$\begin{aligned} y'_j - y_j &= F(\theta + \delta u_j) - F(\theta) + (\varepsilon_j - \varepsilon'_j) = \langle \nabla_\theta F(\theta + \tilde{u}_j), \delta u_j \rangle + (\varepsilon'_j - \varepsilon_j) \\ &= \delta \langle \nabla_\theta F(\theta), u_j \rangle + \delta \langle \nabla_\theta F(\theta + \tilde{u}_j) - \nabla_\theta F(\theta), u_j \rangle + (\varepsilon'_j - \varepsilon_j). \end{aligned}$$

Dividing both sides of the above equality by δ, we obtain

$$\frac{y'_j - y_j}{\delta} = \langle \nabla_\theta F(\theta), u_j \rangle + \underbrace{\langle \nabla_\theta F(\theta + \tilde{u}_j) - \nabla_\theta F(\theta), u_j \rangle}_{:= b_j} + \underbrace{\delta^{-1}(\varepsilon'_j - \varepsilon_j)}_{:= s_j}. \tag{2}$$

Next, define $X = (u_1; u_2; \cdots, u_m) \in \mathbb{R}^{m \times d}$ as an $m \times d$ matrix with each row corresponding to a probing vector u_j; $z, b, s \in \mathbb{R}^n$ as n-dimensional vectors with $z_j = (y'_j - y_j)/\delta$, $b_j = \langle \nabla_\theta F(\theta + \tilde{u}_j) - \nabla_\theta F(\theta), u_j \rangle$, $s_j = (\varepsilon'_j - \varepsilon_j)/\delta$ for $j = 1, 2, \cdots, n$. The estimate $\hat{\nabla}_\theta F(\theta)$ can then be written as $\hat{\nabla}_\theta F(\theta) = (X^\top X)^{-1}(X^\top z)$. In addition, $z = X\nabla_\theta F(\theta) + b + s$ thanks to Eq. (2). Subsequently,

$$\hat{\nabla}_\theta F(\theta) - \nabla_\theta F(\theta) = (X^\top X)^{-1} X^\top (b + s) = (X^\top X)^{-1} X^\top b + (X^\top X)^{-1} X^\top s. \tag{3}$$

Define $\beta := (X^\top X)^{-1} X^\top b$. By Lemma 5, we know that $\|(X^\top X)^{-1}\|_{\mathrm{op}} \leq 2d/m$ with probability $1 - \delta$. Therefore,

$$\|(X^\top X)^{-1} X^\top b\|_2 \leq \|(X^\top X)^{-1}\|_{\mathrm{op}} \|X\|_{\mathrm{op}} \times \sqrt{m}\|b\|_\infty \leq \frac{2d}{m} \times \sqrt{m} \times \sqrt{m} \times \|b\|_\infty \leq 2d\|b\|_\infty,$$

with probability $1 - \delta$, where $\|b\|_\infty = \max_j |b_j|$. Because $\nabla_\theta F(\theta)$ is L-Lipschitz continuous, $|b_j|$ can be upper bounded by

$$\begin{aligned} |b_j| &= \left| \langle \nabla_\theta F(\theta + \tilde{u}_j) - \nabla_\theta F(\theta), u_j \rangle \right| \leq \|\nabla_\theta F(\theta + \tilde{u}_j) - \nabla_\theta F(\theta)\|_2 \times \|u_j\|_2 \\ &\leq L \times \|\tilde{u}_j\|_2 \times \|u_j\|_2 \leq L\delta, \end{aligned}$$

where the last inequality holds because $\|\tilde{u}_j\|_2 = \lambda\delta\|u_j\|_2 \leq \delta$ since $\|u_j\|_2 = 1$ and $0 < \lambda < 1$. Subsequently, we have with probability 1 that

$$\|\beta\|_2 = \|(X^\top X)^{-1}X^\top b\|_2 \leq 2Ld\delta. \tag{4}$$

We next establish a co-variance upper bound on $\zeta = (X^\top X)^{-1}X^\top s$. It should be noted that each $s_j = (\varepsilon_j - \varepsilon'_j)/\delta$ is a centered, independent random variable with variance upper bounded by $\mathbb{E}[s_j^2|\theta] \leq 2/\delta^2$. Subsequently, we have

$$\mathbb{E}[\zeta^\top\zeta|\theta] = \mathbb{E}[s^\top X(X^\top X)^{-2}X^\top s|\theta] = \frac{2}{\delta^2}\mathrm{tr}\left[X(X^\top X)^{-2}X^\top\right] \leq \frac{2}{\delta^2} \times d \times \frac{2d}{m} = \frac{4d^2}{m\delta^2}. \tag{5}$$

Combining Eqs. (4,5) we complete the proof of Lemma 1. □

2.3.3 The ThreadStopping Component

In the ThreadStopping component, we discuss rules for *stopping* a computing thread j if it is deemed to be not promising, either unable or too time-expensive to converge to a good-quality parameter estimate. While the proposed rules are heuristics in nature, we prove under certain local concavity assumptions that these proposed stopping rules are *conservative* in the sense that, with high probability, they will not remove a promising computing thread by mistake.

We describe the two major stopping rules considered in the ThreadStopping component, which can be categorized at a higher level as "first-order" and "second-order" rules.

2.3.3.1 First-order stopping rule Suppose thread j is currently at a parameter estimate θ_j, with gradient estimate $\hat{\nabla}F(\theta_\tau^{(j)}) \approx \nabla F(\theta_\tau^{(j)})$. Suppose also that an estimate $\hat{F}(\theta_\tau^{(j)}) \approx F(\theta_\tau^{(j)})$ is obtained by simply averaging all observations of y_j in Algorithm 2. The thread j should be terminated, then, if there exists another thread $j' \neq j$ such that

$$\textbf{Stopping rule 1:} \qquad \hat{F}(\theta_\tau^{(j)}) + D \times \|\hat{\nabla}F(\theta_\tau^{(j)})\|_2 \leq \hat{F}(\theta_\tau^{(j')}).$$

Here, $D > 0$ is a tuning parameter for the stopping rule and also the optimization algorithm, with larger D values indicating more aggressive (and hence less "safe") stopping rule for computing threads.

Intuitively, the stopping rule obtains an "over-estimate" of the likelihood function F on the solution thread j can potentially converge to, on the left-hand side of the stopping rule. The intuition is that if $\hat{F}(\theta_\tau^{(j)})$ is already small, it means that the function will change very slowly in a neighborhood of $\theta_\tau^{(j)}$ and therefore the estimating gradient ascent procedure is unlikely to advance/improve the likelihood objective significantly in thread j.

Below we state a local concavity condition and shows that, with the condition held and the estimates $\hat{F}, \hat{\nabla} F$ being accurate, the proposed stopping rule is "safe" in the sense that it will only remove computing threads impossible to obtain better parameter estimates. This is further accomplished by showing that, $\hat{F}(\theta_\tau^{(j)}) + D \times \|\hat{\nabla} F(\theta_\tau^{(j)})\|_2$, under the considered circumstances, is an upper bound on how large $F(\theta_j^*)$ could potentially be.

Condition 1 (Local concavity)**.** *For every computing thread j let θ_j^* be the parameter the thread converges to. There exists a convex neighborhood U_j containing θ_j^* with diameter $\sup_{x,x' \in U_j} \|x - x'\|_2 \leq D$, such that F is* concave *in U_j, meaning that*

$$F(\lambda x + (1-\lambda)x') \geq \lambda F(x) + (1-\lambda)F(x'), \qquad \forall x, x' \in U_j, \;\; \lambda \in [0,1].$$

Lemma 2. *Suppose F is twice differentiable, condition 1 holds and $\theta_\tau^{(j)} \in U_j$. Then $F(\theta_j^*) \leq F(\theta_\tau^{(j)}) + D\|\nabla F(\theta_\tau^{(j)})\|_2$.*

Proof. Proof of Lemma 2. Because F is twice differentiable and locally concave on U_j, we know that $\nabla^2 F(\theta) \preceq 0$ for all $\theta \in U_j$. Using second-order Taylor expansion of $F(\theta_j^*)$ at $\theta_\tau^{(j)}$ with Lagrangian remainders, it holds that

$$F(\theta_j^*) = F(\theta_\tau^{(j)}) + \langle \nabla F(\theta_\tau^{(j)}), \theta_j^* - \theta_\tau^{(j)} \rangle + \frac{1}{2}(\theta_j^* - \theta_\tau^{(j)})^\top \nabla^2 F(\tilde{\theta})(\theta_j^* - \hat{\theta}_\tau^{(j)}),$$

where $\tilde{\theta} = \lambda\theta_j^* + (1-\lambda)\theta_\tau^{(j)}$ for some $\lambda \in (0,1)$, and $\tilde{\theta} \in U_j$ since U_j is a convex domain. This implies that $\nabla^2 F(\tilde{\theta}) \preceq 0$ and therefore $(\theta_j^* - \theta_\tau^{(j)})^\top \nabla^2 F(\tilde{\theta})(\theta_j^* - \hat{\theta}_\tau^{(j)}) \leq 0$. Subsequently,

$$F(\theta_j^*) \leq F(\theta_\tau^{(j)}) + \|\nabla F(\theta_\tau^{(j)})\|_2 \times \|\theta_j^* - \theta_\tau^{(j)}\|_2 \leq F(\theta_\tau^{(j)}) + D\|\nabla F(\theta_\tau^{(j)})\|_2,$$

where the first inequality is by the Cauchy-Schwarz inequality and the second inequality holds because $\|\theta_j^* - \theta_\tau^{(j)}\|_2 \leq D$. □

2.3.3.2 Second-order stopping rule The first stopping rule we developed in the previous section could be strengthened if the likelihood objective F has finer properties locally around θ_j^*. In this section we consider a second-order stopping rule, which stops a particular thread j if there exists another thread $j' \neq j$ such that

$$\textbf{Stopping rule 2:} \qquad \hat{F}(\theta_\tau^{(j)}) + \frac{1}{2\alpha} \times \|\hat{\nabla} F(\theta_\tau^{(j)})\|_2^2 \leq \hat{F}(\theta_\tau^{(j')}).$$

Comparing the stopping rule 2 with stopping rule 1, the major difference is the *squared* ℓ_2-norm of the estimated gradients of F at $\hat{\theta}_\tau^{(j)}$. Inuitively speaking, this rule is more "aggressive" than stopping rule 1, since when thread j approaches θ_j^* it converges to, the gradients would be close to zero and therefore $\|\hat{\theta}_\tau^{(j)}\|_2^2$ would be much smaller than $\|\hat{\theta}_\tau^{(j)}\|_2$.

Below we state a local *strong-concavity* condition, which is stronger than Condition 1 for stopping rule 1. We then show, in Lemma 3 below, that under the stronger condition a "clearner" version of stopping rule 2 will not remove computing threads that are still relevant.

Condition 2 (local strong-concavity)**.** *For every computing thread j let θ_j^* be the parameter the thread converges to. There exists a convex neighborhood U_j containing θ_j^* with diameter* $\sup_{x,x' \in U_j} \|x - x'\|_2 \leq D$*, such that F is* α-strongly concave *in U_j, meaning that*

$$\nabla^2 F(\theta) \preceq -\alpha I, \qquad \forall \theta \in U_j.$$

To see why Condition 2 is strong than Condition 1, recall that a twice-differentiable function f is concave on U if $\nabla^2 f(x) \preceq 0$ for all $x \in U$. This is weaker than Condition 2 with some $\alpha > 0$.

Lemma 3. *Suppose F is twice differentiable, condition 2 holds and $\theta_\tau^{(j)} \in U_j$. Then $F(\theta_j^*) \leq F(\theta_\tau^{(j)}) + \frac{1}{2\alpha}\|\nabla F(\theta_\tau^{(j)})\|_2^2$.*

Proof. Proof of Lemma 3. Using second-order Taylor expansion of $F(\theta_j^*)$ at $\theta_\tau^{(j)}$ with Lagrangian remainders, it holds that

$$F(\theta_j^*) = F(\theta_\tau^{(j)}) + \langle \nabla F(\theta_\tau^{(j)}), \theta_j^* - \theta_\tau^{(j)} \rangle + \frac{1}{2}(\theta_j^* - \theta_\tau^{(j)})^\top \nabla^2 F(\tilde{\theta})(\theta_j^* - \hat{\theta}_\tau^{(j)}),$$

where $\tilde{\theta} = \lambda \theta_j^* + (1-\lambda)\theta_\tau^{(j)}$ for some $\lambda \in (0,1)$, and $\tilde{\theta} \in U_j$ since U_j is a convex domain. Because $\nabla^2 F(\widetilde{\theta}) \preceq -\alpha I$ and $|\nabla F(\theta_\tau^{(j)}), \theta_j^* - \theta_\tau^{(j)})| \leq \|\nabla F(\theta_\tau^{(j)})\|_2 \times \|\theta_j^* - \theta_\tau^{(j)}\|_2$ thanks to the Cauchy-Schwarz inequality, we conclude that

$$F(\theta_j^*) \leq F(\theta_\tau^{(j)}) + \|\nabla F(\theta_\tau^{(j)})\|_2 \times \|\theta_j^* - \theta_\tau^{(j)}\|_2 - \frac{\alpha}{2}\|\theta_j^* - \theta_\tau^{(j)}\|_2^2.$$

Completing the squares, we obtain

$$\begin{aligned} F(\theta_j^*) - F(\theta_\tau^{(j)}) &\leq -\frac{\alpha}{2}\|\theta_j^* - \theta_\tau^{(j)}\|_2^2 + \|\nabla F(\theta_\tau^{(j)})\|_2 \times \|\theta_j^* - \theta_\tau^{(j)}\|_2 \\ &= -\frac{\alpha}{2}\left(\|\theta_j^* - \theta_\tau^{(j)}\|_2 - \frac{\|\nabla F(\theta_\tau^{(j)})\|_2}{\alpha}\right)^2 + \frac{\|\nabla F(\theta_\tau^{(j)})\|_2^2}{2\alpha} \\ &\leq \frac{\|\nabla F(\theta_\tau^{(j)})\|_2^2}{2\alpha}. \end{aligned}$$

Re-arranging the terms we have $F(\theta_j^*) - F(\theta_\tau^{(j)}) \leq \frac{\|\nabla F(\theta_\tau^{(j)})\|_2^2}{2\alpha}$, which is to be demonstrated. □

2.4 Numerical Results on Synthetic Data

In this section we report numerical results of our proposed algorithm on the optimization task of a synthetic function. We consider the problem of maximizing a 5-dimensional non-concave function with multiple local maxima and saddle points. To construct such a function, we use the probability-density function of a Gaussian Mixture Model (GMM). More specifically, we consider the following objective function

$$f(x) = \sum_{k=1}^{5} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left\{-\frac{\|x-\mu_k\|_2^2}{2\sigma_k^2}\right\},$$

with $\mu_k = e_k \in \mathbb{R}^5$ being the coordinate basis functions, $\sigma_k = 1.0$ for $k \in \{1,2,3,4\}$ and $\sigma_k = 0.5$ for $k = 0.5$. The construction of the objective function f ensures that it has at least five local minima with similar values, with the local minima tilting towards the last component $k = 5$ being comparably higher due to its smaller variance. More specifically, via accurate calculation, the global maxima of $f(\cdot)$ appears at $x^* = [0.0511, 0.0511, 0.0511, 0.0511, 0.7957]$, very close to the center of the last Gaussian component, with an objective value of $f(x^*) = 1.4572$. Hence, if an (estimating) gradient descent algorithm is initialized near the first for components it will be attracted to the first four local maxima first before escaping and turning towards the final global maxima near the last component.

In Figure 13 we report the convergence of our proposed algorithm with five threads, initialized to solutions close to each of the component centers μ_k defined in the objective function. Both algorithms are run for a total of $T = 100$ gradient evaluations (total number of time periods across all 5 threads), with each gradient evaluation taking $m = 20$ random samples with $\delta = 0.1$ probing radius. In the left panel of Figure 13, both the thread coordination and thread stopping components are disabled (meaning that each time we select a thread uniformly at random, and no thread is terminated early); in the right panel of Figure 13, the thread coordination (thread sampling) component is activated with parameters $\kappa_{\mathsf{CP}} = \kappa_{\mathsf{EP}} = \kappa_{\mathsf{V}} = 1$ with history window $\Delta_t = 100$, and the thread stopping component is activated with rule $\hat{F}(\theta_\tau^{(j)}) + 0.5\|\theta_\tau^{(j)}\|_2 \leq \hat{F}(\theta_\tau^{(j')})$. In both plots of Figure 13, we also report in the dashed blue curve the objective function values of a hypothetical single-thread

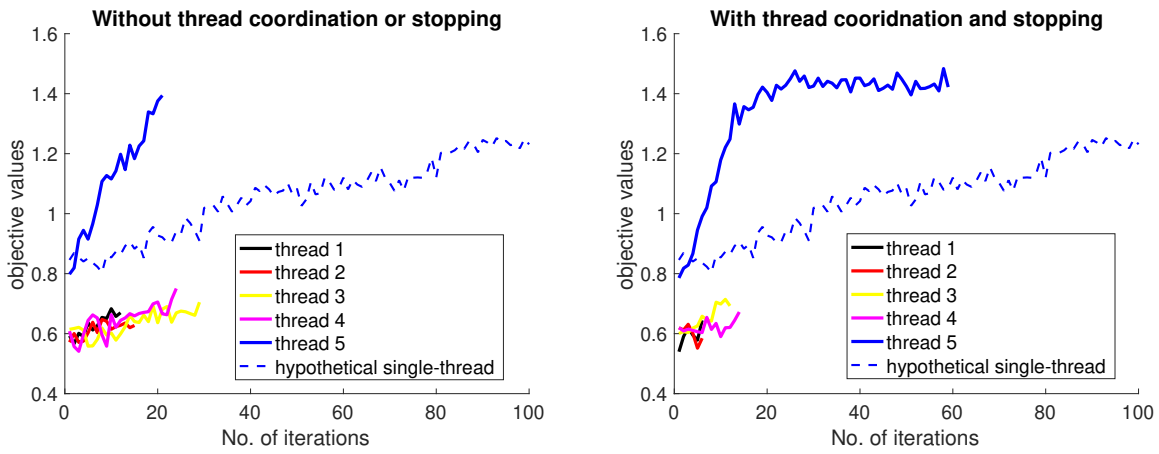


Figure 13: Convergence of our proposed algorithm with five threads. Details of the figures and the algorithms being implemented are given in the main text. Note that the most promising thread #5 converges to the optimal objective $f^* = 1.4572$ rather quickly even with few number of function evaluations.

estimating gradient descent method, initialized at a point close to a sub-optimal component center μ_k, $k < 5$.

From Figure 13, it is clear that our proposed algorithm with multiple threads (more precisely 5 threads in this experiment) outperform its single-thread version significantly, with the same number of function value/gradient evaluations ($T = 100$). Furthermore, the right panel of Figure 13 shows that our thread coordination and thread stopping components could quickly identify sub-optimal threads (those marked with red, yellow, black and magenta curves) and stop them, while the same multi-thread optimization algorithm without thread coordination or stopping is forced to almost evenly distribute the computation among the five threads, wasting computation on unpromising threads and thereby slowing the overall progress of the algorithm.

We also report the global (overall) convergence of the proposed algorithms in Figure 14. For the single-thread curve (the dashed black curve) Figure 14 coincides with Figure 13. For the other two curves corresponding to five threads, we report the best objective value the algorithm attains after *a total of* $t = 1, 2, \cdots, 100$ gradient/function value evaluations

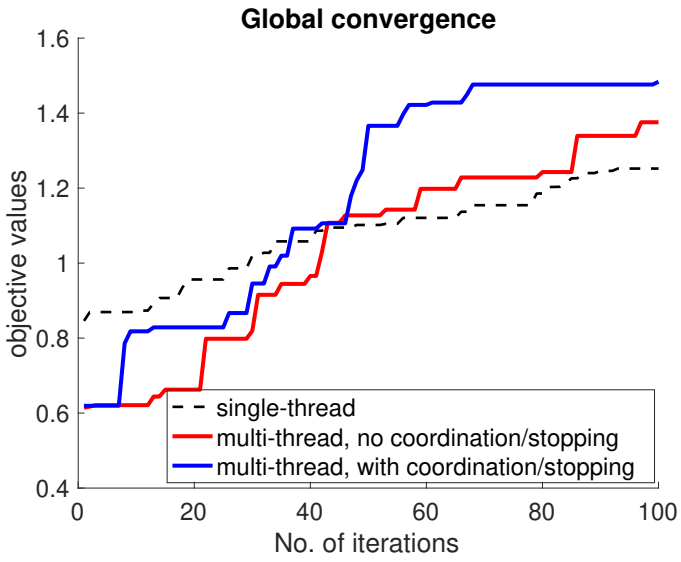


Figure 14: Global (overall) convergence of the proposed algorithm for single-thread, five-thread without coordination/stopping and five-thread with coordination and stopping. Further details are given in the main text.

are made. This gives the reader a more intuitive picture of the efficiency of the proposed optimization algorithms. As we can see, the algorithm with both thread coordination and thread stopping components activated (the solid blue curve) converges to higher objective values much faster than the same algorithm with the same number of threads, but with neither thread coordination nor thread stopping rules (the solid red curve). This is because with thread coordination (sampling the more promising threads more frequently) and thread stopping (terminating sub-optimal threads early to not waste more computation time on these threads), the algorithm allows more samples/computation to be spent on the promising thread so that the convergence speed of the algorithm is much faster.

2.5 Numerical Results on Airline Pricing Data

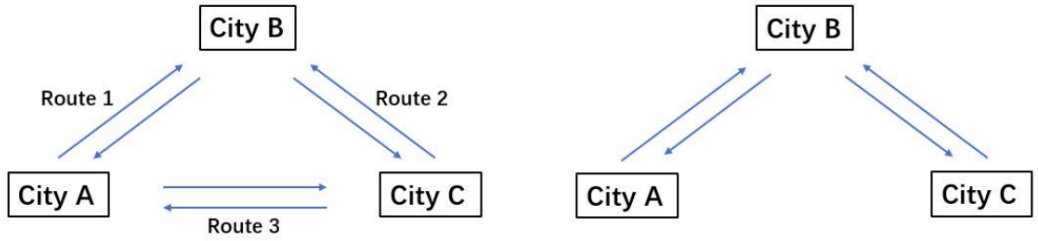
In this section, we apply our proposed optimization method to a real-world airline pricing dataset and compare its performance with benchmark heuristics optimization algorithms, including the *genetic algorithm* and the *pattern search* algorithm. Our proposed algorithm is implemented in C++, while both benchmark methods are implemented in Matlab.

2.5.1 Background: Hidden City Ticketing

Modern airlines operator primarily two types of flight networks: the hub-and-spoke network, which designated a handful of airports as *hubs* and route most of the flights from all airports to major hubs; and the fully-connected network, which operates direct flight in a point-to-point manner. The hub-and-spoke network is adopted by major airlines such as the United airlines and the Delta airlines, while the fully-connected approach is mostly used by smaller, low-cost carriers such as Southwest and JetBlue.

In hub-and-spoke network, many flights between non-hub airports are carried out using connecting flights. For example, the flight from Pittsburgh to Boston could be direct/non-stop, but most likely it needs connection at a New York airport. Naturally, connecting flights are priced (sometimes significantly) lower compared to direct flights due to the additional connection and extended travel time. In some extreme cases, the price of the indirect flight (e.g., Pittsburgh to Boston connecting via New York) might be even lower than the first-leg of the flight (Pittsburgh to New York non-stop). This creates the possibility of *hidden-city ticketing*, a practice that could significantly reduce the revenue/profits of the airlines.

More specifically, consider a traveller who wishes to travel from airport A to airport B. The practice of hidden-city ticketing is defined as the same traveller purchasing a ticket from airport A to airport C, connecting at airport B. The traveller would then proceed with only the first leg of her purchased ticket, essentially traveling from A to B on a non-stop flight. Clearly, such hidden-ticketing behavior is only practiced if the traveller knows about the practice, and furthermore the price of a flight ticket from A to C connecting at B is strictly lower than the price of a non-stop ticket directly from A to B.



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

Figure 15: Illustration of fully-connected (FC) and hub-and-spoke (HS) structures among three airports.

2.5.2 Model Formulation and Maximum Likelihood Estimation

We collected commercial flights operated among the 133 busiest commercial service airports in the United States (identified by the FAA) with 96.34% of total passenger enplanements, 16,142 routes and 2,822,086 itineries.

For every three distinct airport tuples (A, B, C), with flights connecting (A, B) and (A, C), the flight network can be categorized as either *fully-connected (FC)*, if there are non-stop flights between (A, C), or *hub-and-spoke (HS)* if there are no direct flights between A and C. Figure 15 gives a graphical illustration of the FC and/or HS network structures among the airports A, B and C. We use d_{AB}, d_{BC}, d_{AC} to denote the distances between pairs of airports, and p_{AB}, p_{BC}, p_{AC} for ticket prices of *direct* flights among the airports. Additionally, we use p_{ABC} for the ticket price of the indirect flight from A to C connecting at B.

We make the following assumptions on the supply side (the airlines):

1. There is only one airline serving the three cities, thus the firm charges monopoly airfares;
2. 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;
3. Direct flight has a quality of q_h per mile and indirect flight has a quality of q_l per mile, with $0 < q_l < q_h < 1$.

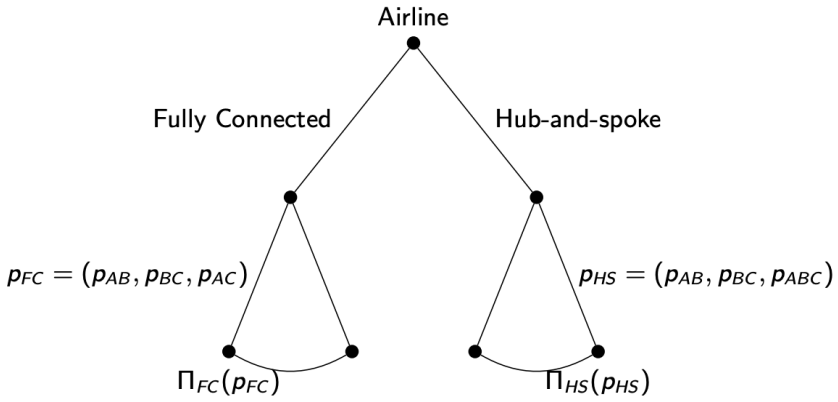


Figure 16: The Stackelberg game description of the airlines' network choices and travelers' riding behaviors.

We also assume the following on the demand side (the travelers):

1. Each individual i has a time preference parameter of λ_i, obtaining utility $C_1 e^{\lambda_i} qd - p$ from consuming a good of quality q, and 0 if he/she does not fly;
2. On each route j, the distribution of consumers' time preferences $\lambda_{ij} \sim \mathcal{N}(\theta_j, \sigma_1^2)$;
3. 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 $1 - \delta$;
4. 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};
5. Amount of passengers on each route j are normalized to 1.

Based on the above assumptions, we use a Stackelberg game to characterize the airlines' network choices and the travelers' riding behaviors, as shown in Figure 16.

We next formulate the maximum likelihood estimation question that we aim to solve. The problem is centered on the parameter of interest δ, representing the portion of travelers who are aware and would be willing to exploit the benefits of hidden city ticketing. Other nuisance parameters also needed to be estimated include C_1 (the travelers' average utility), C_2 (the airlines' average cost), σ_1 (standard deviation of λ_{ij}), σ_2 (standard deviation of

$\theta_j \sim \mathcal{N}(\mu_j, \sigma_2^2)$), q_h and q_l (utility multipliers for different types of travelers). Mathematically, we use

$$\zeta = (\delta, C_1, C_2, \sigma_1, \sigma_2, q_h, q_l) \in \mathbb{R}^7$$

to denote a vector of parameter values.

For each airports tuple (A_i, B_i, C_i) in the collected data such that there are at least one flights between (A_i, B_i) and (B_i, C_i), respectively, we use

$$y_i \in \{0, 1\}$$

to denote the airline's actual choice of FC or HS structures (i.e., $y_i = 1$ if there are direct flights between (A_i, C_i) and $y_i = 0$ otherwise). The accessible data for airport tuple (A_i, B_i, C_i) are represented as

$$x_i = (p_{AB}, p_{BC}, p_{AC}, p_{ABC}, d_{AB}, d_{BC}, d_{AC}, \mu_B, \mu_C) \in \mathbb{R}^9,$$

where $p_{AB}, p_{BC}, p_{AC}, p_{ABC}$ are flight ticket prices, d_{AB}, d_{BC}, d_{AC} are distances between airports, and μ_B, μ_C are expected values of θ_B, θ_C which are obtained via analyzing the typical flows of flights/passengers for each of the 133 airports in the data set.

With $\theta_B \sim \mathcal{N}(\mu_B, \sigma_2^2)$ and $\theta_C \sim \mathcal{N}(\mu_C, \sigma_2^2)$ realized, the expected profit for the airline can be computed as follows. Let $\Phi(z; \mu, \sigma^2) = \int_0^z \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(t-\mu)^2}{2\sigma^2}\} \mathrm{d}t$ be the cumulative density function (CDF) of $\mathcal{N}(\mu, \sigma^2)$. First, in the case of Fully-Connected (FC) network, the expected profit is

$$\Pi_{FC} = \Pi_{AB} + \Pi_{BC} + \Pi_{AC} \text{ where } \Pi_{XY} = p_{XY} \left[1 - \Phi\left(\ln\left(\frac{p_{XY}}{C_1 q_h d_{XY}}\right); \theta_Y, \sigma_1^2\right)\right] - C_2 d_{XY}.$$

In the case of Hub-and-Spoke (HS) network, there are two cases. The first case is $p_{AB} < p_{ABC}$, in which there is no hidden-city ticketing opportunities. The airline's profit is then expressed as

$$\Pi_{HS} = \Pi_{AB} + \Pi_{BC} + R^l_{ABC} \text{ where } R^l_{ABC} = p_{ABC} \left[1 - \Phi\left(\ln\left(\frac{p_{ABC}}{C_1 q_l (d_{AB} + d_{AC})}\right); \theta_C, \sigma_1^2\right)\right].$$

Note that, because the direct flight between A and C is no longer operated in a hub-and-spoke network, and therefore R^l_{ABC} is a pure profiting term.

Finally, in the case of $p_{AB} \geq p_{ABC}$, there is potential profit loss due to hidden-city ticketing. The airline's profit is

$$\Pi_{HS} = (1-\delta)\Pi_{AB} + \Pi_{BC} + R^l_{ABC} + \delta R^h_{ABC}$$
$$\text{where } R^h_{ABC} = p_{ABC}\left[1 - \Phi\left(\ln\left(\frac{p_{ABC}}{C_1 q_h d_{AB}}\right); \theta_B, \sigma_1^2\right)\right],$$

where $\delta \in (0,1)$ is the parameter of interest corresponding to the portion of travelers engaged in the hidden-city ticketing practice.

Given tuple x_i and the observed airline's network choice $y_i \in \{0,1\}$, the log-likelihood of y_i conditioned on x_i and parameter ζ can be written as

$$\log P(y_i|x_i;\zeta) = y_i \log \Pr_{\theta_{B_i},\theta_{C_i}}\left[\Pi_{FC} > \Pi_{HS}\right] + (1-y_i)\log \Pr_{\theta_{B_i},\theta_{C_i}}\left[\Pi_{HS} \geq \Pi_{FC}\right]. \tag{6}$$

The maximum-likelihood estimation problem is then formulated as

$$\arg\max_{\zeta} \frac{1}{N}\sum_{i=1}^{N} \log P(y_i|x_i;\theta), \tag{7}$$

where N is the total number of airport tuples (A,B,C) available in the data collected.

2.5.3 Results

Before presenting the computational results, we first mention some important implementation details. First, because the distances d_{XY} are measure in miles and could vary drastically, we adopt a re-normalization transform $d_{XY} \mapsto \sqrt{d_{XY}}$ to alleviate the scales of the distances. We also note that the log-likelihood in Eq. (6) cannot be evaluated directly because the $\Pr_{\theta_{B_i},\theta_{C_i}}\left[\Pi_{FC} > \Pi_{HS}\right]$ terms do not admit easy closed-form expression. Instead, we use Monte-Carlo sampling with M_{MC} samples to approximately compute the log-likelihood. Finally, because the collected data consist of many airport tuples (i.e., $N > 10^5$), we use a "mini-batch" approach when evaluating the objective function in Eq. (7). More specifically, we randomly sample $M_{\text{MB}} \ll N$ airport tuples and use the average log-likelihood on the randomly sampled mini-batch data to approximate the log-likelihood of the objective function on the entire data set.

Table 5: Results for our proposed algorithm on the airline pricing data, with $D = 1/2\alpha \in \{1.0, 0.5, 0.3\}$. A $\times$ means that the particular thread is not active at the end of the optimization.

	$D = 1/2\alpha$	log. likeli.	δ	C_1	C_2	σ_1	σ_2	q_h	q_l	running time (s)
Thread #1	1.0	-0.69	0.01	6.33	0.76	0.69	0.82	0.53	0.10	108.9
	0.5	-0.69	0.02	6.34	0.78	0.72	0.85	0.52	0.10	107.7
	0.3	-0.68	0.02	6.33	0.78	0.72	0.87	0.52	0.10	108.3
Thread #2	1.0	-0.73	0.49	11.8	0.95	0.35	0.86	0.54	0.12	108.9
	0.5	-0.71	0.50	11.8	0.94	0.35	0.93	0.52	0.12	107.7
	0.3	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
Thread #3	1.0	-0.71	0.06	10.3	0.74	0.37	0.96	0.49	0.10	108.9
	0.5	-0.70	0.07	10.3	0.74	0.39	0.99	0.47	0.11	107.7
	0.3	-0.69	0.07	10.3	0.76	0.43	1.0	0.47	0.11	108.3
Thread #4	1.0	-0.68	0.50	10.9	0.95	0.72	0.83	0.34	0.10	108.9
	0.5	-0.69	0.49	10.9	0.96	0.72	0.82	0.33	0.12	107.7
	0.3	-0.67	0.48	10.9	0.91	0.73	0.86	0.33	0.12	108.3
Thread #5	1.0	-0.71	0.43	12.0	0.96	0.59	0.77	0.36	0.11	108.9
	0.5	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
	0.3	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

Table 6: Results for the genetic algorithm (`ga`) and the pattern search algorithm (`ps`). Each algorithm terminates only when the designated time limit is reached. M_{MB} equals to M_{MC} are the mini-batch sizes and the number of Monte-Carlo samples, respectively.

	M_{MB}	M_{MC}	log. likeli	δ	C_1	C_2	σ_1	σ_2	q_h	q_l	running time (s)
Genetic algorithm	200	200	-0.73	0.30	5.03	0.68	0.61	0.86	0.85	0.27	100
	200	200	-0.75	0.28	5.36	0.73	0.58	0.78	0.82	0.28	300
	500	500	-0.69	0.31	3.66	0.70	0.81	0.95	0.93	0.21	600
Pattern search	200	200	-0.82	0.13	7.5	0.10	0.10	0.85	1.0	0.10	100
	200	200	-1.01	0.0	8.0	0.10	0.10	0.60	1.0	0.10	300
	500	500	-0.76	0.35	0.34	0.35	0.51	0.76	0.58	0.16	600

We also impose the following constraints on each of the parameters in ζ to be estimated: $\delta \in (0, 1/2]$, $C_1 \in [1, 20]$, $C_2 \in [0.1, 1]$, $\sigma_1 \in [0.1, 1]$, $\sigma_2 \in [0.1, 1]$ and $0.1 \leq q_l \leq q_h \leq 1$. These constraints are imposed to ensure the values of the unknown parameters are practically feasible and reasonable.

We first present computational results for our proposed parallel stochastic gradient ascent algorithm. Algorithmic parameters are set as $\kappa_{\mathsf{CP}} = \kappa_{\mathsf{EP}} = \kappa_{\mathsf{V}} = 1$ for the ThreadCoordination component, $m = 20$, $\delta = 0.05$ for the GradientEstimation component, and $D = 1/2\alpha \in \{1.0, 0.5, 0.3\}$ for the ThreadStopping component. The algorithm is run for $T = 200$ iterations, 5 initial threads with random initializations, and with the log-likelihood of the final solution of each active threads being reported in Tables **??**.

As we can see from Table **??**, at the end of the optimization there are 5 active threads if $D = 1/2\alpha = 1.0$, 4 active threads if $D = 1/2\alpha = 0.5$ and 3 active threads if $D = 1/2\alpha = 0.3$. This is intuitive because smaller $D = 1/2\alpha$ values indicate a more aggressive thread elimination policy, thereby leading to fewer active threads. In the core three threads (Threads #1, 3, 4) the log-likelihood is consistently below -0.7, indicating likely scenarios in the real world. Our algorithm takes roughly 100 seconds to complete an optimization.

We next compare the results obtained by our proposed algorithm with two baseline methods widely used in global optimization: the *genetic algorithm* and the *pattern search*

algorithm. Both algorithms have been implemented in Matlab in standard packages, via the `ga` and the `patternsearch` routine. The results for both algorithms are reported in Table 6, with their corresponding running times. Note that the running times are set prior to each run, and the optimization algorithms simply terminate once the time budgets are reached. Table 6 shows that, both the genetic algorithm and the pattern search algorithm take much longer time (around or even more than 10 minutes) to converge to less optimal solutions compared to the ones found by our proposed algorithm in 100 seconds. This demonstrates the effectiveness and efficiency of our proposed approach.

2.6 Conclusion

In this paper, we propose a general framework of optimizing functions with noisy function evaluations. The proposed framework is based on running multiple threads of stochastic estimating gradient ascent algorithms in parallel, and to carefully coordinate the different computing threads. Theoretical analysis and justifications are given for the stopping rules used in the proposed method, and numerical results on both synthetic data and a real airline industry data set are provided to corroborate the effectiveness and efficiency of our proposed methods.

3.0 Does H-1B Visa Reforms Affect Whether US Natives Major in STEM Fields?

This chapter exploits large changes in the H-1B visa program and examines the effect of changes in H-1B admission levels on the likelihood that US natives major in STEM fields. Compare to effect on labor market outcomes, the possible impact of H-1B visa reforms on natives' college major choices indicate effect over longer horizons. I find some evidence that H-1B population adversely affect natives' choices in STEM fields when they enter the college and graduate from it. Female, male and White subgroups have been negatively affected, and the native Asian subgroup suffer from the most dramatic crowd-out effect. Given that the H-1B population share had been more than doubled during 1992 to 2017, the probability of native Asian graduates majoring in STEM fields would be 2.56 percentage points larger, if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Since foreign-born Asian account for a large proportion of H-1B visa holders, there might be an interesting "Asian crowd out Asian" story.

3.1 Introduction

The number of college students major in STEM (science, technology, engineering and mathematics) fields is commonly viewed as critical to the long-term technology advancement and economic growth of United States. Nowadays, there has been concern that not enough US natives are studying in STEM fields, and one possible reason is that they might be crowded out by foreign-born students. According to the data released by 2009-17 American Community Survey (ACS), the proportion of US natives major in STEM fields varied within the range of 15% to 25% during year 1960 to 2017. It was relatively stable fluctuating around 20% in 1990s and 2000s, and kept increasing after 2010. But the percentage of foreign-born students major in STEM fields has been always higher than that of native graduates, and it showed a nearly 15% increase during 1960 to 2017. Theoretically, large amount of foreign-

born students studying STEM majors might be a double-edged sword for native graduates. Foreign-born students could possibly crowd out natives of STEM majors because they are competing for limited education resources, or there might be positive spill-over effect instead when natives are attracted and retained in those fields.

The relationship between natives' college major choices and H-1B visa reforms might be trivial at first glance. But according to the data released by U.S. Citizenship and Immigration Services (USCIS), large proportion of H-1B visa holders work in STEM related occupations. Thus the H-1B visa program governs most admissions of foreign-born graduates with a bachelor's degree or above major in STEM fields for employment in United States. Therefore, whether the large changes in this program will affect US natives' major choices in STEM fields or not becomes an interesting research question to explore.

Since it was created in year 1992, the controversy over H-1B visa program never stops. Proponents emphasize that those high-skilled workers are important to the technology advancement of US economy, and if H-1B visa is contracted, "America is losing many very skilled workers ... They are losing their dreams, and America is losing the value they bring". ([86]). While the detractors keep worried about native US workers being displaced by foreign-born workers, or furthermore, native US students being crowded out by foreign-born students in especially STEM fields. Compare to the impact on natives' labor market outcomes, possibility of affecting the major choices in STEM fields of native students might be even more worthy of studying, because the effect could be on the US economy operation over longer horizons.

To bring identification to the research question, following [48], I will exploit large changes in the H-1B population over the 1992-2017 period. The H-1B population fluctuated substantially during this period because firstly, the national cap on new H-1B issuance varied a lot from a lower bound of 65,000 new workers a year to a higher bound of 195,000. Secondly, the usage of cap and total H-1B issuance also varied a lot due to the change of policy and economic condition. According to the summary statistics published by USCIS, large proportion of H-1B specialty occupation workers are young (between the ages of 25 and 34), well educated (with a bachelor's degree or above), working in STEM-related occupations and earn relatively high median salary.

This study focuses on the relationship between natives' college major choices and H-1B visa reforms. More specifically, I am trying to measure the impact of changes in H-1B population on the probability that US native students choose to major in STEM related fields when they enter the college and graduate from it. I choose the undergraduate level education because this is the key step for an individual to obtain a STEM degree and work in STEM related occupations after graduation. The ACS is a large-scale survey conducted by the U.S. Census Bureau every year and it has asked respondents with at least a bachelor's degree to report their college majors since 2009. Although there are previous literatures looking at the outcomes of H-1B visa program, and literatures focusing on the possible factors that affect students' major choices, to the best of my knowledge, this is the first paper examining the relationship between H-1B visa reforms and US natives' college major choices in STEM fields directly.

More specifically, this paper measures the possible impact of changes in H-1B admission levels on the likelihood that US natives major in STEM fields over the 1992-2017 period. To bring identification of this problem, I exploit the variation of H-1B population shares across areas and over time. This is not easy due to data limitation. Exploiting the variation of H-1B population across more narrowly defined labor markets is difficult with standard data resources. Therefore, I have applied a innovative approach exploiting the micro-level data in the first step of H-1B visa application. Another challenge is to identify the causal relationship, which is also difficult because of the endogeneity of immigrants' self location choices. Hence I have applied an instrumental variable approach constructing an IV based on the historic settlement pattern of foreign-born STEM workers. To better interpret the results, I also take a further look at the impacts on different gender and race subgroups. This is a reasonable approach because different subgroups vary from one to another. According to the data, much less female students choose STEM majors compare to male students, while much more Asian natives choose to major in STEM fields compare to other race groups. Existing literatures also find some heterogeneity among different gender and race subgroups. For example, [43] shows that the persistence in STEM majors is much lower for women and minorities. [64] also find similar results, while [69] find the opposite. Analyzing the results for different subgroups could reveal a more complete story and help shed light on

policy implications. For all of my preferred specifications, I have included control variables of lagged H-1B population, personal characteristics, and labor market conditions. State and year fixed effects and state-specific linear trends are also included to make the causal relationship more valid.

In this paper, I have found significant negative impacts of H-1B population shares on natives majoring in STEM fields, both when they enter the college and graduate from it. For students beginning their college education, a 10% increase in the H-1B population share decreases the probability of native students choosing STEM majors by 0.032%, decreases the likelihood of a male native student majoring in STEM related fields by 0.07%, and decreases that of a White native student by 0.025%. Given that the H-1B population share had been more than doubled from year 1992 to 2017, the probability of native students majoring in STEM fields when they enter the college would be 0.74 percentage points larger, if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Similarly, the likelihood of male native students choosing STEM majors would be 1.62 percentage points larger, and that of White native students would be 0.58 percentage points larger. For students graduating from college, results indicate that a 10% increase in the H-1B population share decreases the probability of native graduates choosing STEM majors by 0.021%, decreases the likelihood of a female native graduate majoring in STEM related fields by 0.014%, decreases that of a male native graduate by 0.032%, and decrease that of a White native graduate by 0.02%. The native Asian subgroup suffer from the most dramatic crowd-out effect. A 10% increase in the H-1B population share would decrease the likelihood of an Asian native graduate majoring in STEM related fields by as large as 0.111%. Again, given that the H-1B population share had been more than doubled, the probability of native graduates majoring in STEM fields when they graduate would be 0.49 percentage points larger, the likelihood of female native graduates choosing STEM majors would be 0.32 percentage points larger, that of male native graduates would be 0.74 percentage points larger, and that of White native graduates would be 0.46 percentage points larger. For the native Asian subgroups, the probability would have been as large as 2.56 percentage points larger if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Since foreign-born Asian account for a large proportion of H-1B

visa holders, there might be an interesting "Asian crowd out Asian" story here.

The remainder of this chapter is organized as follows: The introduction of related works is summarized in Section 3.2. The background information including the US college major choices and the H-1B visa program, as well as the details of the dataset being used are explained in more details in Section 3.3 and Section 3.4 respectively. Section 3.5 describes the empirical framework in details, including the plain Probit regression model and the instrumental variable approach. Section 3.6 shows the results of both approaches and Section 3.7 concludes.

3.2 Literature Review

Previous literatures mainly focus on the labor market outcomes of H-1B visa program. [55], [57], [50], and [70] provide some general information about H-1B visa program, H-1B population estimates and characteristics of H-1B visa holders. The evidence of impact of H-1Bs on labor market outcomes are mixed. Some papers find that the H-1B visa holders adversely affect native workers' employment opportunities, wages, etc. For example, [56] raise some concern given trends in the postdoctoral labor market and for employers in 'job shops' who undercut US workers with temporary workers. [61] criticize that the industry's motivation for hiring H-1Bs is primarily a desire for cheap, compliant labor, and show the adverse impacts of the H-1B program on various segments of the American computer-related labor force. [50] find some evidence of aggressive wage-cost cutting, including paying H-1B recipients only the legally mandated 95 percent of the prevailing US wage, among some H-1B employers. In contrast, some other papers show positive impacts of H-1B workers on natives' earnings, employment rate, etc. For example, [87] find some positive relationship between LCAs (Labor Condition Applications, the first step towards H-1B visa application) and earnings, earnings growth, and the unemployment rate in the IT sector at the state level. [48] show that higher H-1B admissions increase immigrant science and engineering (SE) employment and patenting by inventors with Indian and Chinese names in cities and firms dependent upon the program relative to their peers. [44] find that immigrants who

entered on a temporary work visa have a large advantage over natives in wages, patenting, and publishing, and are more likely to start companies than similar natives. [65] show that increases in STEM workers are associated with significant wage gains for college-educated natives, and foreign STEM increased total factor productivity growth in US cities.

Besides the mixed evidence of impacts of H-1B workers on native workers' labor market outcomes, there are also a few studies paying attention to the H-1B visa program on educational outcomes. For example, [47] find that restrictive immigration policy disproportionately discourages high-ability international students from pursuing education in the United States. And [4] show that the binding cap of H-1B visa raises international students' likelihood of employment in academia, even outside of their field of study.

With respect to students' college major choices, existing literatures have revealed different factors that might contribute to this decision making process - in STEM or non-STEM fields - and whether students tend to persist or change their majors during college, using both reduced-form and structural model approaches. For example, [22] find that the number of foreign PhD students in sciences majors shows a positive effect on undergraduate students also choosing sciences majors. [7] show that academic background can fully account for average differences in switching behavior between blacks and whites. [58] find that weaker, non-minority students typically respond to greater competition in the sciences by shifting their major choice. [63] find some evidence that immigration adversely affects whether US-born women who graduated from college majored in a science or engineering field. [6] show significant sorting into majors based on academic preparation, with science majors at each campus having on average stronger credentials than their non-science counterparts. [9] find that students with relatively greater non-STEM ability are more likely to switch out of STEM. [5] estimate a dynamic model of the ability sorting across majors and conclude that virtually all ability sorting is because of preferences for particular majors in college and the workplace. [8] estimate a model of college major choice that incorporates subjective expectations and assessments and show that both expected earnings and students' abilities in the different majors are important determinants of a student's choice of a college major. [3] develop a dynamic model of educational decision-making and figure out the important role for heterogeneity in tastes for fields of study and the occupations they lead to.

To summarize, there are previous literatures looking at the outcomes of H-1B visa program, and literatures focusing on the possible factors that affect students' major choices, while to the best of my knowledge, this is the first paper to exam the relationship between H-1B visa reforms and US natives' college major choices in STEM fields directly.

3.3 U.S. College Major Choices

My paper applied data on college majors from the 2009-17 ACS. The ACS is a large-scale survey conducted by the U.S. Census Bureau every year. Since 2009, it started to ask respondents with at least a bachelor's degree to report their college major. To define STEM majors, the Department of Homeland Security (DHS) has made a STEM Designated Degree Program list, which is a complete list of fields of study that DHS considers to be science, technology, engineering or mathematics (STEM) fields of study for purposes of the 24-month STEM optional practical training extension. According to the regulation, a STEM field of study is a field of study "included in the Department of Education's Classification of Instructional Programs taxonomy within the two-digit series containing engineering, biological sciences, mathematics, and physical sciences, or a related field. In general, related fields will include fields involving research, innovation, or development of new technologies using engineering, mathematics, computer science, or natural sciences (including physical, biological, and agricultural sciences)". Combine the STEM Designated Degree Program list with the field of degree information provided by ACS, I have established a list of STEM majors in appendices Table 13. Compare to the classification in [63], which includes only majors in biology and life sciences, physical sciences, engineering, computer and information sciences, and mathematics and statistics as STEM majors, my classification is broader and fits the STEM optional practical training extension well, which is highly correlated with the H-1B visa program. This broader classification has increased the percentage of STEM majors in my raw data from 19.74% (according to the definition of STEM majors in [63]) to 22.64%.

Assume that the traditional college age is aged 18-22. Figure 17 shows the percentage of US native college graduates who majored in STEM fields when they were age 22, the

modal age when they graduated, during year 1960 to 2017, as well as that of foreign-born students. The solid line indicates that the proportion of US natives majoring in STEM fields varied within the range of 15% to 25% during the period. It declined in the 1960s and rose in the 1970s. The increase in the 1970s and early 1980s might reflect the emphasis on science and math during that time ([63]). A sharp decrease occurred in mid-1980s and the ratio went back to around 20% until 2010. It seems that the Internet boom of the late 1990s did not have a significant impact on native students' major choices in STEM fields. The ratio kept increasing after year 2010. Due to the data collecting process, foreign-born students being included in the analysis are those who were living in the United States when the ACS was conducted. The dot line shows that a much higher percentage of foreign-born students choosing to major in STEM related fields compared to native graduates, and the ratio revealed a nearly 15% increase during this period. After the introduction of H-1B visa program since early 1990s, the ratio kept growing when the program expanded and decreased when the policy contracted in early 2000s. After that it began rebounding when the macroeconomy started to recover from the crisis, and the H-1B population also kept growing since then.

Figure 18 shows the percentage of US native college graduates majoring in STEM fields by gender group. In general, the proportion of male and female STEM students share similar trends during this time period. The proportion of male students choosing STEM majors is much higher than that of female students, whereas female students showed a more stable increasing trend compared to that of the male students. Figure 19 shows the percentage of US native college graduates majoring in STEM fields by race group. I have divided the native students into four different race groups: non-Hispanic white, non-Hispanic black, Asian and Hispanic. The non-Hispanic other race is not shown. The solid line representing the non-Hispanic white group is smooth with largest number of observations. Dot line for non-Hispanic black group and dash line for Hispanic group show smaller percentages of native students majoring in STEM fields, compared to that of the non-Hispanic white group. The Asian group, represented by the dash dot line, indicates a much higher and more volatile proportion of STEM students compared to all the other three groups.

In my empirical analysis, I will match the state-level data of H-1B population with

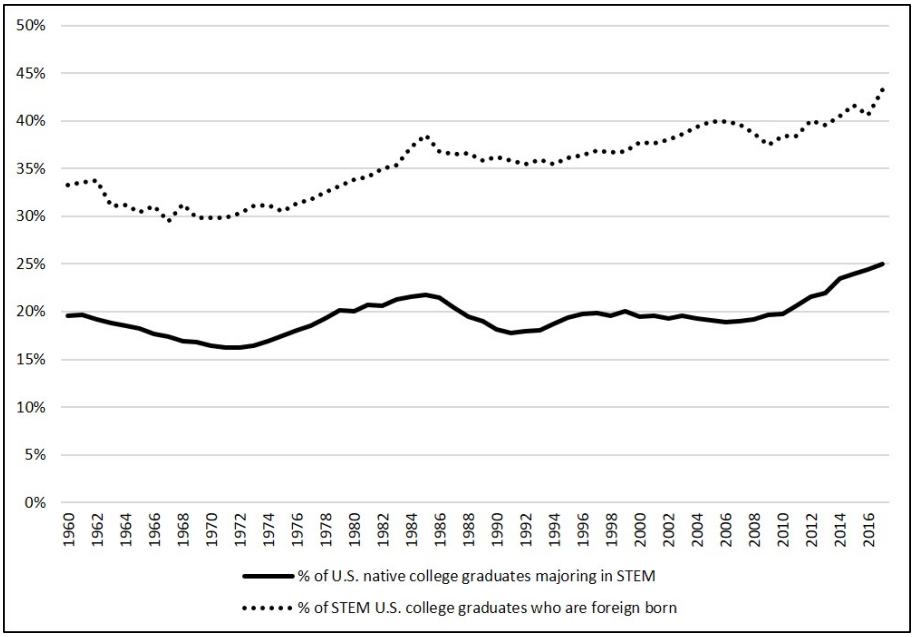


Figure 17: The proportion of college graduates majoring in STEM by nativity.

individuals' state of birth. State of birth is the only place of residence available in the ACS besides the current place and place one year ago, and state of birth is also highly correlated with the state of having college education in United States, compared to the other two places. There are also other previous literatures using the state of birth to examine state-level variables related to college education. ([29], [35], [21], [63]) An advantage of using state of birth is to mitigate the endogeneity selection bias that might arise if the H-1B population affects US native college students' location choice of college attendance. According to the data published by National Center for Education Statistics, the average ratio of all first-time degree/certificate-seeking undergraduates in degree-granting postsecondary institutions entering a college within their state of birth was as high as 81.80% in year 1992 (Utah has the highest proportion of 94% and Connecticut has the lowest of 59%, with District of Columbia being excluded) and 78.38% in year 2016 (Utah has the highest proportion of 90.85% and Vermont has the lowest of 50.60%, with District of Columbia being excluded). Figure 20 shows the variation of proportion of in-state students across states when they entered the

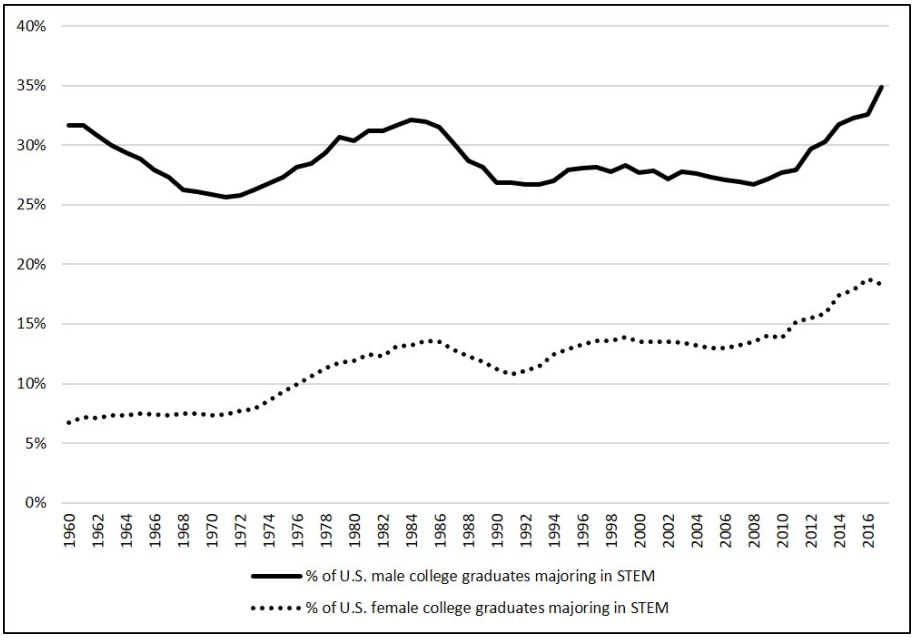


Figure 18: The proportion of U.S. native college graduates majoring in STEM by gender.

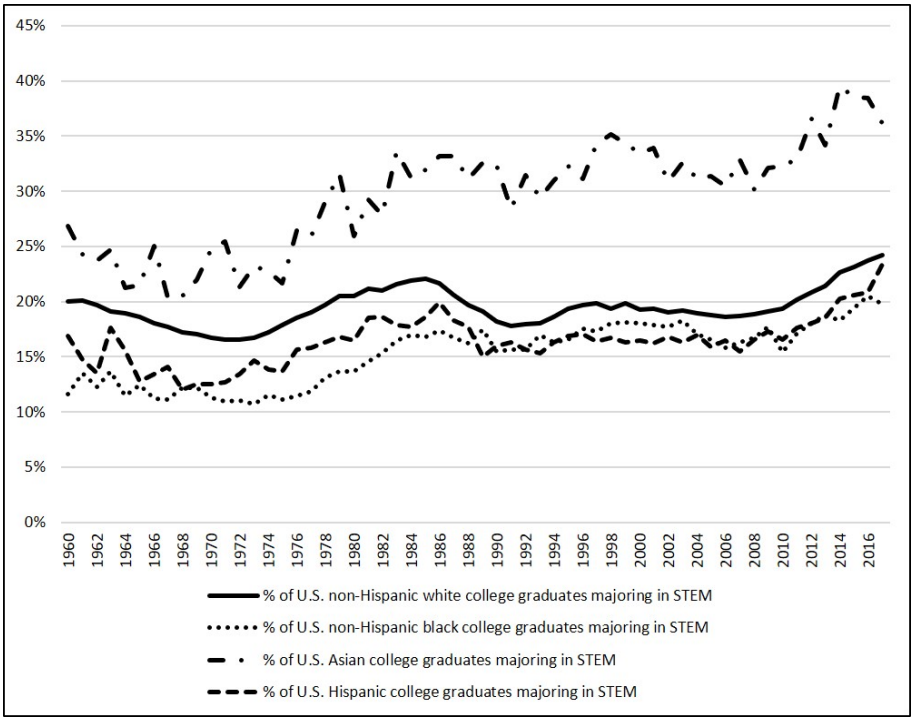


Figure 19: The proportion of U.S. native college graduates majoring in STEM by race.

college in year 1992 and 2016, respectively. From the literature and the data, we should be able to reasonably conclude that state of birth is a good indicator of the state where US natives have their college education. Since not all of the students entering college in their state of birth, my estimates are likely to underestimate the possible impact of H-1B visa reform on US natives' college major choices, and this underestimation would be slightly different across states.

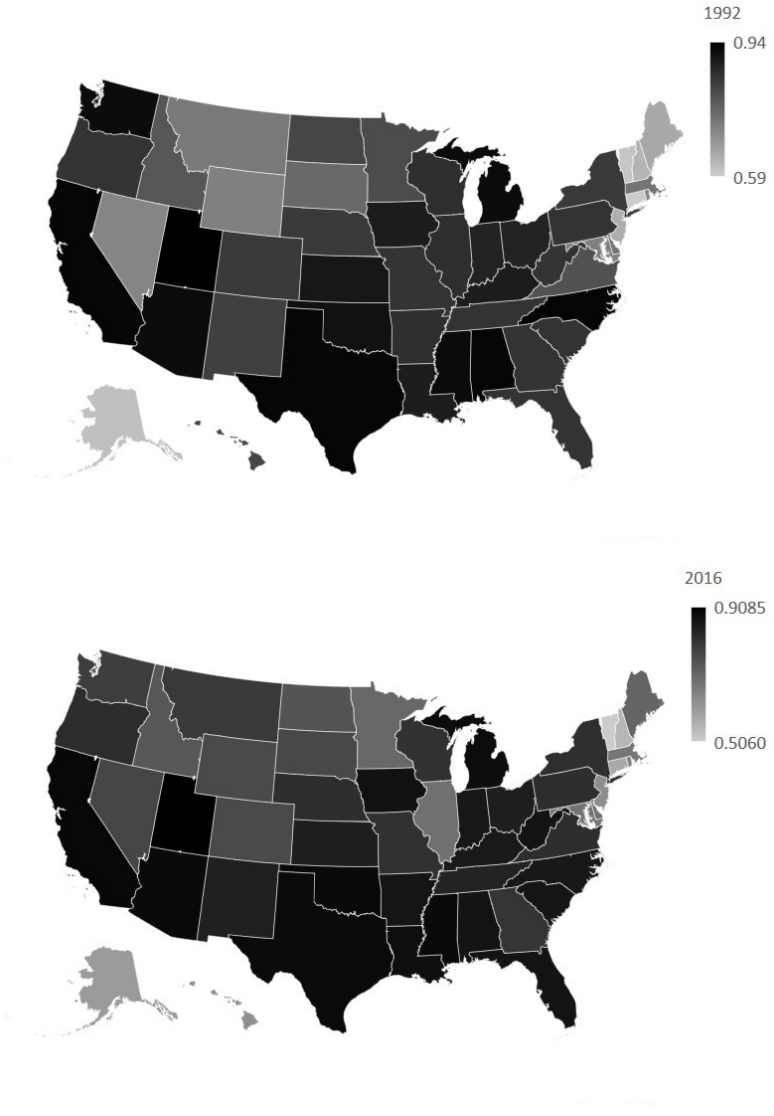


Figure 20: The variation of proportion of U.S. natives entering college within their state of birth across states (year 1992 and 2016, District of Columbia excluded).

3.4 H-1B Visa Program

The H-1B is a visa in the United States that allows U.S. employers to temporarily employ foreign workers in specialty occupations. The regulations define a specialty occupation as "requiring theoretical and practical application of a body of highly specialized knowledge in a field of human endeavor including but not limited to biotechnology, chemistry, computing, architecture, engineering, statistics, physical sciences, medicine and health, and requiring the attainment of a bachelor's degree or its equivalent as a minimum" (U.S. Code Title 8, Chapter 12, Subchapter II, Part II, Section 1184 - Admission of nonimmigrants).

Since year 2004, USCIS started to publish the Annual Report to Congress including Report on H-1B Petitions and Characteristics of H-1B Specialty Occupation Workers, both of which provide some summary statistics of H-1B workers. In general, H-1B specialty occupation workers are young (between the ages of 25 and 34) and well educated (with a bachelor's degree or above). More than 80% of those visa holders work in STEM related occupations, such as occupations in computer sciences, engineering, mathematics, physical sciences and life sciences. They also earn much higher median salary compared to the U.S. average. For example, in year 2017, the median salary of beneficiaries of approved H-1B petitions increased to $85,000, much higher than the nominal median income per capita of $31,786 and the real median household income of $61,372 reported by the Census Bureau. More details of the characteristics of H-1B workers are summarized in Table 7.

Because the H-1B visa is a necessity for a foreign-born graduate to work in United States, any reform of it towards to or away from STEM students would reasonably affect foreign-born students' major choices in especially STEM fields, and therefore, might crowd US natives out or have positive spillovers on them through attracting or retaining them in those fields.

The H-1B visa program has been changing ever since it started. The Immigration Act of 1990 (implemented in 1992) created the H-1B visa for professional foreign nationals seeking temporary employment in the United States. At the time of its creation, 65,000 H-1B visas became available for new applicants each year. The cap was not reached until fiscal year 1997 and again in 1998. In October 1998, Congress enacted the American Competitiveness

Table 7: Characteristics of H-1B specialty occupation workers.

Year	**% age 25-34**	**% bachelor's degree**	**% master's degree**	**% doctorate degree**	**% computer-related occupations**	**Median salary**
2004	65.5	48.7	33.9	11.2	44.5	$53,000
2005	65.6	44.8	36.8	5.3	43.0	$55,000
2006	66.1	45.0	38.6	10.6	48.4	$60,000
2007	65.7	44.0	40.4	10.1	49.8	$60,000
2008	66.1	43.0	40.6	10.9	49.6	$60,000
2009	65.9	40.9	39.9	12.6	41.6	$64,000
2010	67.7	42.5	39.2	11.7	47.5	$68,000
2011	69.7	41.4	41.7	5.2	50.8	$70,000
2012	72.1	46.3	40.8	8.4	59.5	$70,000
2013	70.9	45.0	41.1	8.8	59.8	$72,000
2014	71.7	45.0	43.1	7.8	64.5	$75,000
2015	71.0	45.0	44.0	7.0	66.5	$79,000
2016	68.9	44.2	45.4	6.9	69.1	$82,000
2017	66.2	45.2	44.5	6.8	69.8	$85,000

and Workforce Improvement Act (ACWIA), which temporarily raised the cap to 115,000 for fiscal years 1999 and 2000. Limit were both reached. Congress responded to the increase in demand for H-1B visas with the American Competitiveness in the 21st Century Act (AC21). The act had two relevant effects. First, it reduced the number of H-1B visas that counted toward the quota by exempting employees of universities, nonprofit research organizations, and governmental research organizations. Second, it raised the cap to 195,000 for each of year 2001, 2002, and 2003. Those limits were never reached. In year 2004, Bachelor's degree cap returned to 65,000 with added 20,000 visas for applicants with U.S. postgraduate degrees. The H-1B cap has been binding every year since then. On April 2, 2008, the U.S. Department of Homeland Security announced a 17-month extension to the OPT for students in qualifying STEM fields. And the 17-month extension has been replaced by a longer 24-month extension since May 10, 2016, which allows the foreign STEM students to work up to 36 months under their student visa, and provides them as long as three years to obtain an H-1B visa.

Figure 21 shows the annual H-1B visa issurance cap (dash line), the H-1B visa issuance for initial employment (dot line), and the H-1B visa population estimate (solid line) during the fiscal year 1992 to 2017. In FY 1999 the actual issuance exceeded the national cap because of a computer malfunction announced by Immigration and Naturalization Service (INS). In recent years, the decoupling of the actual issuances and the numerical cap is due to the policy change that employees working for specific institutions have been exempted from the quota.

The solid line in Figure 21 represents an important estimation of H-1B population from FY 1992 to 2017. In this paper, I use the population stock rather than the net issuances or the cap as my primary explanatory variable. Because the change of H-1B population stock reflects not only the change of H-1B inflow but also the outflow, it provides more complete information when we want to see the possible impact of H-1B visa reforms on natives' college major choices. Estimating the H-1B visa population is not straightforward. Although the initial H-1B visa issuance provides a reasonably good measurement of the inflow, the outflow estimation needs to be modeled carefully. [55] provides one way to model the outflow of H-1B pool using the information of transitions to permanent residency, emigration, and death.

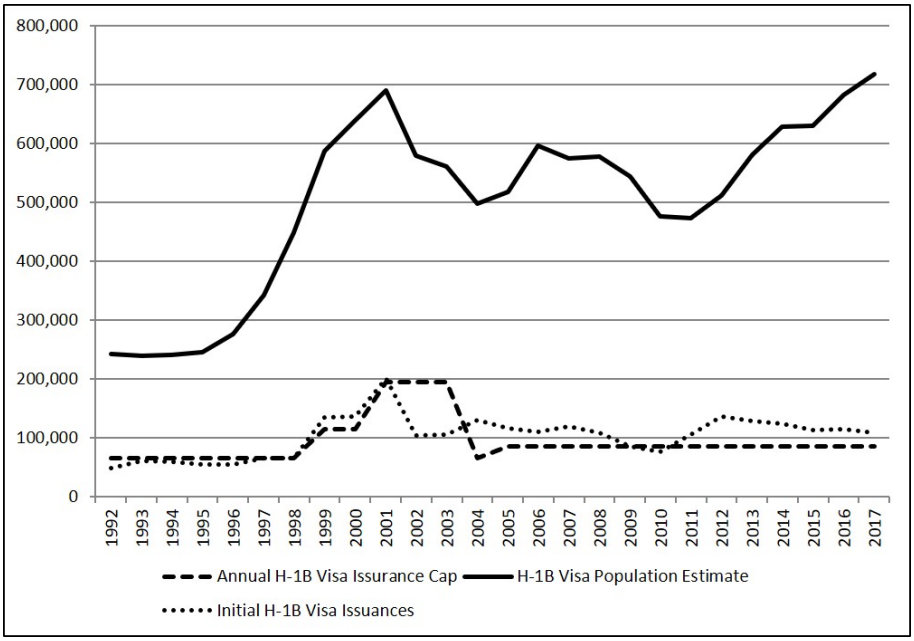


Figure 21: H-1B visa issuance and population estimates.

And [48] applies Lowell's updated estimates in their 2010 paper. In this study, I will measure the outflow in a more innovative and precise way. Following [33], I assume that the outflow of H-1B pool results not only from people adjusting to lawful permanent resident (LPR) status, emigration and death, but also from some measurement error because of H-1B one-year extensions, possible duplicate petitions, and people changing for employers. Data of adjustment to LPR status come from the yearbook of INS and DHS. Emigration estimations come from [14] paper. I explore the CDC's (Centers for Disease Control and Prevention) online tool to compute an estimated H-1B mortality rate, as that of Asian and Pacific Islander males between the ages of 25 and 34 by the year 2017. Information of changing employers are from the USCIS yearbook. And according to [33], estimated rate of duplicate petitions approved is around 1% and appoved H-1B one-year extensions estimated by DHS is 18.3%. In Figure 21 we can see that the resulting change in H-1B visa population is large enough to be economically important.

Beyond those nation-level broad statistics, research related to H-1B visa program has

been largely restricted by the data limitation. In order to exploit the variation across more narrowly defined labor markets, and control for the many contemporaneous national changes occurring within the United States during the same period, I follow the methodology in [48] and exploit the more micro trends using the data of Labor Condition Applications (LCAs) published by U.S. Department of Labor. According to the regulation, for every H-1B petition filed with the USCIS, a LCA must first be certified by the U.S. Department of Labor to ensure that the wage offered to the non-immigrant worker meets or exceeds the "prevailing wage" in the area of employment, so that U.S. workers' wages or working conditions will not be displaced or adversely affected by the foreign workers. A big advantage of LCA data is that it provides much more detailed information of the potential H-1B visa holders, including their work city, county (since 2015) and state since 2001. Although the LCA approvals do not translate one for one into H-1B grants because of the national cap of the H-1B visa issuances, it should be one of the (best) indicators available to shed light on the variation of H-1B population across states.

Since the ACS data only provides state-level information of natives' birthplace, in this paper I will use state as the primary labor market to quantify the possible impact of changing H-1B population on US natives' college major choices in STEM related fields. In order to estimate $\text{H-1B}_{s,t}$, I assume that the variation of H-1B population across states is proportional to the portion of LCAs across states, i.e., $\text{H-1B}_{s,t}$ can be estimated by the following formula,

$$\text{H-1B}_{s,t} = \text{H-1B}_t \cdot \frac{\text{LCA}_{s,t}}{\text{LCA}_t}.$$

Data of $\text{LCA}_{s,t}$ is available since year 2001, thus for year 1992 to 2000, I further assume that the $\text{LCA}_{s,t=1992-2000}$ is equal to the average of $\text{LCA}_{s,t=2001-2017}$.

For every state, I compute the share of H-1B population with respect to state population and rank them according to their "dependency" on H-1B population. Table 8 shows the top 10 and bottom 10 states in year 1992 and 2017 respectively. We can see that there is large variation of H-1B population share across states, ranging from 0.01% (Montana) to 0.34% (District of Columbia, or 0.24% of New Jersey if District of Columbia is excluded) in 1992 and 0.02% (Wyoming) to 0.57% (District of Columbia, or 0.47% of New Jersey if District of Columbia is excluded) in 2017. The standard deviation was as high as 0.06% in 1992

and raised to 0.13% in 2017. The change of H-1B population proportions from 1992 to 2017 reveals that for most of the states, this ratio has been increasing during the time period, with the most dependent states being Washington, California, District of Columbia, New Jersey, and Pennsylvania. Figure 22 also shows the variation of share of H-1B Population across states in year 1992, 2017, and changes from 1992 to 2017 respectively, from the top dependent state to the bottom.

Table 8: Top 10 dependent states on H-1B population in 1992 and 2017.

	1992		2017		change (1992-2017)	
			Top 10 Dependent States			
1	District of Columbia	0.3358%	District of Columbia	0.5732%	Washington	0.3083%
2	New Jersey	0.2425%	New Jersey	0.4658%	California	0.2800%
3	Delaware	0.2249%	Washington	0.4452%	District of Columbia	0.2374%
4	Massachusetts	0.1736%	California	0.4222%	New Jersey	0.2233%
5	Connecticut	0.1572%	Delaware	0.3668%	Pennsylvania	0.1780%
6	California	0.1422%	Massachusetts	0.3511%	Massachusetts	0.1776%
7	Washington	0.1370%	Connecticut	0.2886%	Illinois	0.1682%
8	New York	0.1228%	Illinois	0.2795%	New York	0.1422%
9	Texas	0.1200%	New York	0.2650%	Delaware	0.1419%
10	Georgia	0.1142%	Pennsylvania	0.2617%	Rhode Island	0.1402%
			Bottom 10 Dependent States			
42	Hawaii	0.0269%	Oklahoma	0.0459%	Maine	0.0198%
43	New Mexico	0.0260%	South Dakota	0.0412%	Louisiana	0.0179%
44	Oklahoma	0.0246%	Louisiana	0.0377%	West Virginia	0.0154%
45	Alaska	0.0204%	Alabama	0.0316%	South Dakota	0.0125%
46	Alabama	0.0202%	Montana	0.0308%	Alabama	0.0114%
47	Louisiana	0.0198%	Hawaii	0.0280%	Mississippi	0.0041%
48	Mississippi	0.0166%	West Virginia	0.0277%	Wyoming	0.0026%
49	Wyoming	0.0154%	Mississippi	0.0207%	Hawaii	0.0011%
50	West Virginia	0.0123%	Alaska	0.0196%	Nevada	-0.0002%
51	Montana	0.0110%	Wyoming	0.0180%	Alaska	-0.0007%

3.5 Empirical Framework

In this paper, I use probit regression model to estimate the possible impact of H-1B visa reforms on US natives' college major choices in STEM related fields. The estimating

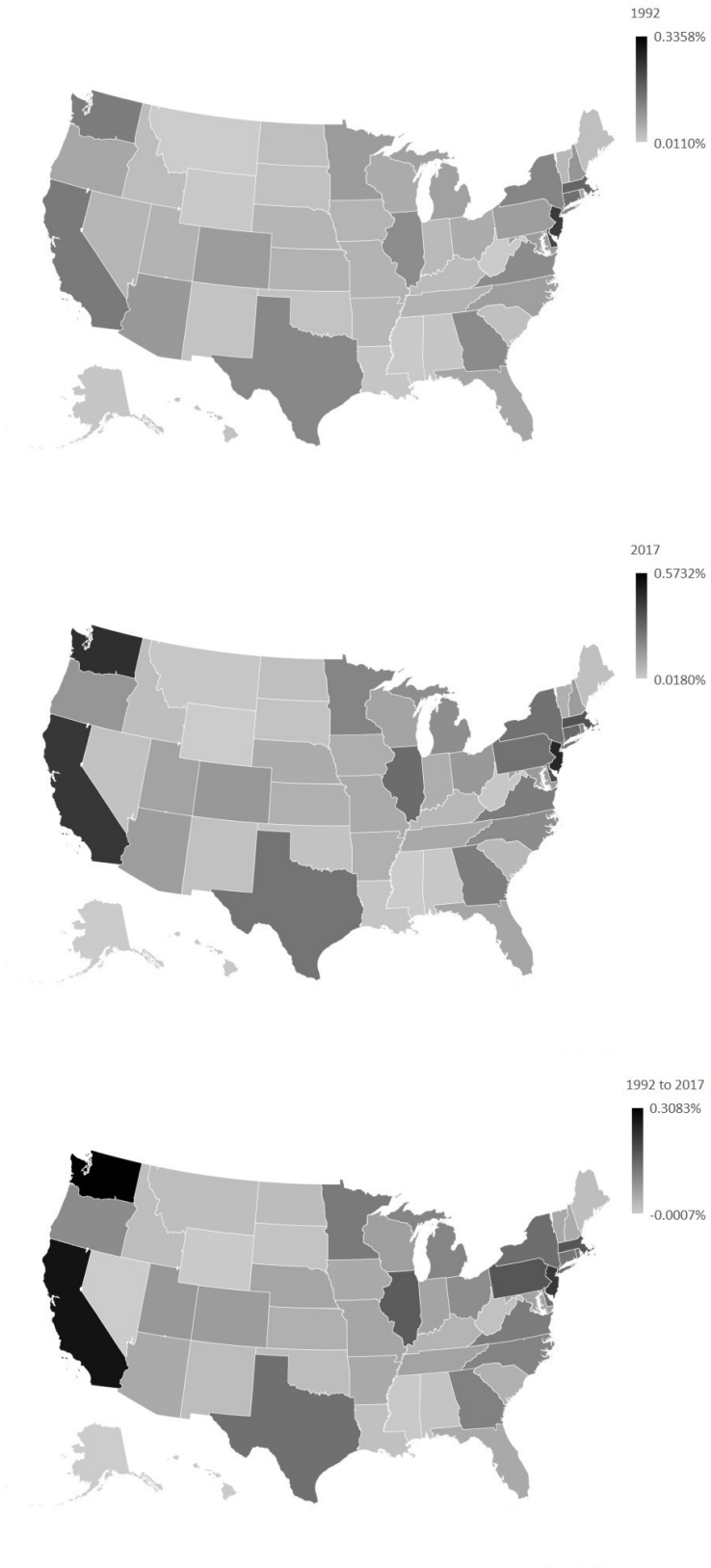


Figure 22: Distribution of share of H-1B population across states in 1992 and 2017.

framework is

$$\text{STEM}_{ist} = I(\alpha + \beta \cdot \ln(\text{H-1B Share}_{st}) + \delta X_{ist} + \theta Z_{st} + \phi_s + \eta_t + \varepsilon_{ist} \geq 0),\ \varepsilon \sim (0, 1).$$

Assume that traditionally students enter their college in age 18 and graduate in age 22. The dependent variable equals 1 if individual i who was born in state s majored in STEM fields when he or she was 18 or 22 years old in year t. H-1B $\text{Share}_{s,t}$ is the estimated H-1B population in state s and year t as a percentage of the state population in state s and year t. X_{ist} represent the characteristics of individual i, including his or her age, age square, gender, and dummy variables for race (white, black, Hispanic, Asian or other). When these variables are included, they are controlling for any systematic differences in the probability of majoring in STEM fields across different sub-groups. Lagged H-1B population share are also included. For year t when individual was 18 years old, I include lagged H-1B population share for the past 3 years, which should traditionally cover the individual's high school education. Under those specifications, I am trying to measure the possible impact of H-1B visa reforms on US natives' college major choices when they enter the college. And for year t when individual was 22 years old, I include lagged H-1B population share for the past 5 years, which are supposed to cover the individual's college education. Therefore, under those specifications, I am trying to measure the possible impact of H-1B visa reforms on US natives' college major choices when they graduate from college, which is a joint choice of major when they enter and shift during the college.

Intuitively, the condition of STEM jobs market could also affect natives' college major choices in STEM fields. Therefore, Z_{st} represent some specifications controlling for the STEM labor market conditions. I use six different measures to control for the relative attractiveness of STEM jobs, including variables that equal the proportion of college graduates working in STEM occupations in state s and year t, the change of that proportion during the past decade, the ratio of total personal income of college graduates being employed in STEM occupations to that of college graduates in non-STEM occupations in state s and year t, the change of that ratio in the past 10 years, the ratio of wage and salary income of college graduates being employed in STEM occupations to that of college graduates in non-STEM occupations in state s and year t, and the change of that ratio during the past decade. STEM

occupations are defined based on the classifications in [65] paper, data source include 1980, 1990, 2000 census, and 2001 to 2017 ACS. Since the corresponding occupation codes change from dataset to dataset, more details are provided in the Appendices Table 14.

The regression model also includes state and year fixed effects, and state-specific linear time trends. The state-of-birth fixed effects control for any unobservable factors that are specific to the state but constant over time, such as climate and location. The year fixed effects control for any unobservable factors that are specific to that year, such as the macroeconomy condition and policy change. The state-specific linear time trend help control for any unobservable, smooth changes within the state that might affect the likelihood that US natives' college major choices in STEM fields. The standard errors ε_{ist} are robust and clustered on the state.

3.5.1 Instrumental Variable Approach

The distribution of H-1B population across states might be endogenous, suffering from the problem of self selection. Factors that affect the possibility of H-1B visa holders to live in one certain state might also affect US natives' college major choices in STEM related fields. If those factors could not be completely captured by my labor market control variables, state and year fixed effects, and state-specific linear time trend, the plain Probit regression estimates will have an upward or downward bias. For example, an upward bias might accur if the H-1B visa holders are attracted by a state with educational systems putting more emphasis on STEM education and imposing policies to generate more STEM majors. On the other side, if there are native families who work in STEM occupations and expect their children to also disproportionately major in STEM majors, and decide to move away from the state with a large number of H-1B population before their children are born, this will end up with a downward bias for the plain Probit estimates.

To deal with this potential endogeneity problem, I will apply an instrumental variable approach besides the plain Probit regression model. The instrument is based on the foreign-born STEM workers' historical settlement patterns. Previous studies have found out that immigrants tend to settle in the same areas as earlier immigrants from their country of origin.

Existing research using instruments for the immigrant share based on historical settlement patterns include but are not limit to [28], [27], [72], [45], [75], etc. This is also well known as the Bartik or shift-share instrument. Two recent papers by [42] and [46] look into the details of this instrument, and provide some econometric and economic reasons to reconsider the conditions under which it could be used. They emphasize that the initial shares need to be exogenous, which is reasonable here because year 1980 was more than ten years before the H-1B visa program getting started. According to the summary statistics of the H-1B visa program, we can see that large proportion of H-1B visa holders are working in STEM related occupations. Therefore, the historical settlement patterns of foreign-born STEM workers before the H-1B visa program started might be a good instrument for the distribution of H-1B population nowadays. My instrument variable is constructed by reallocating the H-1B population across states based on the foreign-born STEM workers' distribution, for 17 countries or regions of origin, across states in 1980. The 17 countries or regions are Africa, India, China, Philippines, South Korea, Japan, Rest of Asia, United Kingdom, Germany, France, Rest of Europe, Canada, Mexico, Rest of North America, Oceania, Brazil, and Rest of South America. These 17 groups are chosen because according to the data provided by Department of State (since 1997) and USCIS (since 2003), each group accounted for a significant share of H-1B visa holders, and each group also accounted for a relatively large share of foreign-born STEM workers in 1980. Therefore, this classification could avoid having large numbers of zeros in my constructed instrumental variables. Recall that year 1980 was more than ten years before the H-1B visa program getting started. Specifically, I compute the predicted H-1B population share in state s and year t according to the following formula,

$$\text{Predicted H-1B Share}_{st} = \frac{\sum_{j=1}^{n} \text{H-1B}_t^j \times \%\text{ of foreign born STEM workers in } s_{1980}^j}{\text{Population}_{st}}.$$

where j represents country or region of origin.

The underlying assumption for this instrument variable to be valid is that the distribution of foreign-born STEM workers by country or region of origin across states in year 1980 is not correlated with any factor that affect US natives' college major choices in STEM related fields occurring more than ten years later. In other words, shocks that affect the distribution of foreign-born STEM workers in 1980 and US natives' college major choices do not persist

over time. This should be a reasonable assumption because year 1980 predates much of the beginning of the H-1B visa program.

3.6 Results

Table 9 and Table 10 show the plain Probit regression results for the correlation between H-1B population shares and US natives' college major choices in STEM fields. And Table 11 and Table 12 report the instrumental variables results. For all the results tables, control variables for personal characteristics and labor market conditions are included, as well as state-of-birth fixed effect, year fixed effect, and state-specific linear trend. Besides the all sample regression, I have also divided the full sample into different sub-groups according to gender and race (White, Black, Hispanic and Asian), in order to see the possible impact of H-1B visa reforms on college major choices of different subgroups. The totals for White, Black, Hispanic and Asian do not sum to the full sample size because otherrace category is not included. As explained in 3.5, lagged H-1B population shares are also included as control variables. For Table 9 and Table 11, I am trying to measure the possible impact of H-1B visa reforms on US natives' major choices when they enter the college. Therefore, I am looking at the coefficients when individuals are in their age 18, which is assumed to be the traditional year of entering college. Lagged H-1B population shares for the past 3 years are included, which normally cover students' high school education. For Table 10 and Table 12, I am trying to measure the possible impact of H-1B visa reforms on US natives' major choices when they graduate from college. Therefore, I am looking at the coefficients when individuals are in their age 22, which is assumed to be the traditional year of graduation. The major choices observed when they graduate are supposed to be a joint choice of majors when they enter and make possible shift during the college. Lagged H-1B population shares for the past 5 years are included, which usually cover students' college education.

Recall that in a linear regression model, we could directly interpreting the estimated coefficients as the marginal effects. But this is not the case for a probit regression model. In general, we cannot interpret the coefficients from the output of a probit regression model

in a standard way. The marginal effects of the regressors refer to how much the conditional probability of the outcome variable changes when we change the value of a regressor, holding all other regressors constant at some values. In particular, the marginal effects depend not only on the regression coefficients, but also on the values of all the other regressors. Therefore, in the probit regression model, there is an additional step of computation required to get the marginal effects once we have computed the probit regression fit. In the result tables below, I have listed both the outputs of the coefficients and the average marginal effects (both with corresponding standard deviations) for the purpose of interpretation. Although these two magnitude are different, their sign and significance level are definitely the same.

3.6.1 Plain Probit Results

From Table 9 we can see that for the full sample, there is no significant relationship found between H-1B population shares and natives' college major choices in STEM fields when they enter the college at age 18. However, if I divide the full sample into different gender and race subgroups, we can see that there are positive effects showing up for the female and the White subgroups, and negative effect for the Hispanic subgroup. In other words, given the choices made when they enter the college, the likelihood of female and White native students choosing STEM majors have been positively affected by the H-1B population shares, while the Hispanic subgroup have been adversely affected.

Further computation of the average marginal effects provide the information of interpreting the magnitude of the impacts. For example, in my Probit regression model with the log of H-1B population share as the explanatory variable, an estimated average marginal effect of 0.0038 suggests that a 10% increase in the H-1B population share increases the probability of a female native student majoring in STEM fields when she enters the college by 0.038 percentage points. Similarly, an estimated average marginal effect of 0.0045 indicates that a 10% increase in the H-1B population share will raise the likelihood of a White native student choosing STEM majors when entering the college by as much as 0.045 percentage points. And an estimated average marginal effect of -0.0212 suggests that a 10% increase in the H-1B population share decreases the probability of a Hispanic native student majoring

in STEM fields by 0.212%.

Table 10 shows the plain Probit regression results when students graduate from college at age 22. We can see that no significant result has been found for the full sample or any gender and race subgroup. Both the crowd-in and crowd-out effects found in subgroups above when students enter the college have disappeared.

3.6.2 Instrumental Variables Results

Given the analysis in Section 3.5, we know that endogeneity of self-selection problem might bias the plain Probit results positively or negatively. Therefore, a more preferred specification would be the instrumental variable approach. Besides showing the regression results of the IV approach of the probit model, both Table 11 and Table 12 also display the F-test statistics from the first-stage of each IV regression. All the F-test statistics are well above 10, which indicate that the instrument variable is valid and has a strong first-stage.

Table 11 reports the IV regression results when natives enter the college at their age 18. We can see that the IV estimates are more negative compared to the plain Probit results, which indicate that the coefficients of the plain Probit regression model might suffer from upward bias. From the result table we can see that for the full sample, the H-1B population shares adversely affects natives' college major choices in STEM fields significantly at the 99% confidence interval. A 10% increase in the H-1B population shares decreases the probability of native students majoring in STEM fields when they enter the college by as large as 0.032 percentage points. If we take a further look at the impacts on different subgroups, we can see that both male and the White subgroups have been negatively affected. A 10% increase in the H-1B population share would decrease the likelihood of a male native student choosing STEM majors by 0.07%, and decrease that of a White native student by 0.025%.

The implied marginal effects might be trivial at first glance. But recall that the H-1B population share had increased dramatically during year 1992 to 2017, from 0.0954% in 1992 to 0.2204% in 2017. Given that it had more than doubled during the period, the probability of native students majoring in STEM fields when they enter the college would be 0.74 percentage points larger - a nontrivial difference - if the H-1B population shares had

remained at their 1992 levels and all else had remained the same. Similarly, the likelihood of male native students choosing STEM majors would be 1.62 percentage points larger, and that of White native students would be 0.58 percentage points larger.

The crowd-out effect remains for the whole sample as well as the male and White subgroups when natives graduate, and even the female and Asian subgroups are adversely affected now. Compare to their major choices when entering the college, students might have a better understanding of the job market they are facing and the career path they are planning when graduation. Therefore, students might choose to change their college majors by shifting to other fields during college education. The final decisions observed in the data show negative estimates for both the full sample and different subgroups. From Table 12 we can see that for the full sample, the H-1B population share adversely affects natives' college major choices in STEM fields significantly when they graduate. Referring to the magnitude, a 10% increase in the H-1B population share decreases the probability of natives choosing STEM fields when graduate by as large as 0.021 percentage points. When we take a further look at the possible impacts on different gender and race subgroups, we can see that more subgroups have been negatively affected. Female, male, White and Asian subgroups have all been adversely affected by the H-1B population share. A 10% increase in the H-1B population share could decrease the likelihood of a female native graduate majoring in STEM fields by 0.014%, decrease that of a male native graduate by a more significant impact of 0.032%, and decrease that of a White native graduate by 0.02%. The native Asian subgroup suffer from the most dramatic negative effect. The parameter estimate indicates that a 10% increase in the H-1B population share would decrease the likelihood of an Asian native graduate choosing STEM majors by as large as 0.111%.

Similarly, given that the H-1B population share had been more than doubled during year 1992 to 2017, the probability of native students majoring in STEM fields when they graduate from college would be 0.49 percentage points larger, if the H-1B population shares had remained at their 1992 levels and all else had remained the same. The likelihood of female native students choosing STEM majors when graduation would be 0.32 percentage points larger, that of male native students would be 0.74 percentage points larger, and that of White native students would be 0.46 percentage points larger. For the native Asian

subgroups, the probability would have been as large as 2.56 percentage points larger if the H-1B population shares had remained at their 1992 levels and all else had remained the same. This is a dramatic difference, and since foreign-born Asian account for a large proportion of H-1B visa holders, there might be an interesting "Asian crowd out Asian" story here.

Table 9: Probit regression estimates for relationship between majoring in STEM and H-1B population share when entering college.

% of H-1B population in:	All	Female	Male	White	Black	Hispanic	Asian
Coefficients:							
Age 18	0.0092	0.0171**	0.0004	0.0166**	0.0197	-0.0854**	0.0178
	(0.0070)	(0.0087)	(0.0111)	(0.0070)	(0.0360)	(0.0380)	(0.0310)
Age 17	0.0016	0.0071	-0.0053	0.0012	0.0248	-0.0060	-0.0345
	(0.0091)	(0.0105)	(0.0122)	(0.0096)	(0.0330)	(0.0367)	(0.0483)
Age 16	-0.0092	-0.0078	-0.0112	-0.0070	-0.0143	-0.0194	-0.0419
	(0.0081)	(0.0114)	(0.0120)	(0.0101)	(0.0397)	(0.0454)	(0.0404)
Age 15	-0.0016	-0.0106	0.0064	-0.0021	0.0543*	-0.0317	0.0224
	(0.0088)	(0.0098)	(0.0145)	(0.0093)	(0.0314)	(0.0375)	(0.0406)
Average marginal effects:							
Age 18	0.0025	0.0038**	0.0001	0.0045**	0.0050	-0.0212**	0.0064
	(0.0019)	(0.0019)	(0.0038)	(0.0019)	(0.0091)	(0.0094)	(0.0111)
Age 17	0.0004	0.0016	-0.0018	0.0003	0.0063	-0.0015	-0.0123
	(0.0025)	(0.0024)	(0.0041)	(0.0026)	(0.0083)	(0.0091)	(0.0173)
Age 16	-0.0025	-0.0017	-0.0038	-0.0019	-0.0036	-0.0048	-0.0150
	(0.0022)	(0.0025)	(0.0041)	(0.0027)	(0.0100)	(0.0112)	(0.0145)
Age 15	-0.0004	-0.0024	0.0022	-0.0006	0.0137*	-0.0078	0.0080
	(0.0024)	(0.0022)	(0.0049)	(0.0025)	(0.0079)	(0.0093)	(0.0145)
# of obs	1,373,608	784,689	588,919	1,101,135	85,070	94,601	56,343

Table 10: Probit regression for relationship between majoring in STEM and H-1B population share when graduation.

% of H-1B population in:	All	Female	Male	White	Black	Hispanic	Asian
Coefficients:							
Age 22	-0.0127	-0.0097	-0.0163	-0.0134	0.0344	-0.0290	-0.0569
	(0.0097)	(0.0096)	(0.0129)	(0.0089)	(0.0265)	(0.0350)	(0.0469)
Age 21	-0.0093	-0.0022	-0.0163	-0.0091	-0.0538**	0.0231	0.0200
	(0.0082)	(0.0098)	(0.0109)	(0.0089)	(0.0259)	(0.0303)	(0.0444)
Age 20	0.0056	0.0025	0.0086	0.0079	0.0320	-0.0603	-0.0131
	(0.0090)	(0.0104)	(0.0116)	(0.0100)	(0.0359)	(0.0376)	(0.0361)
Age 19	-0.0174**	-0.0157	-0.0192*	-0.0187**	-0.0134	0.0264	-0.0499*
	(0.0087)	(0.0106)	(0.0106)	(0.0090)	(0.0397)	(0.0339)	(0.0292)
Age 18	0.0091	0.0176*	-0.0008	0.0155**	0.0309	-0.0886***	0.0186
	(0.0066)	(0.0095)	(0.0109)	(0.0071)	(0.0332)	(0.0313)	(0.0355)
Average marginal effects:							
Age 22	-0.0035	-0.0022	-0.0055	-0.0036	0.0087	-0.0072	-0.0203
	(0.0026)	(0.0021)	(0.0044)	(0.0024)	(0.0067)	(0.0086)	(0.0168)
Age 21	-0.0025	-0.0005	-0.0055	-0.0025	-0.0136**	0.0057	0.0072
	(0.0022)	(0.0022)	(0.0037)	(0.0024)	(0.0065)	(0.0075)	(0.0159)
Age 20	0.0015	0.0005	0.0029	0.0022	0.0081	-0.0149	-0.0047
	(0.0025)	(0.0023)	(0.0039)	(0.0027)	(0.0091)	(0.0093)	(0.0129)
Age 19	-0.0047**	-0.0035	-0.0065*	-0.0051**	-0.0034	0.0065	-0.0178*
	(0.0024)	(0.0024)	(0.0036)	(0.0024)	(0.0100)	(0.0084)	(0.0104)
Age 18	0.0025	0.0039*	-0.0003	0.0042**	0.0078	-0.0218***	0.0067
	(0.0018)	(0.0021)	(0.0037)	(0.0019)	(0.0084)	(0.0077)	(0.0127)
# of obs	1,627,098	925,293	701,805	1,312,411	102,196	108,507	61,990

Table 11: IV regression for relationship between majoring in STEM and H-1B population share when entering college.

% of H-1B population in:	All	Female	Male	White	Black	Hispanic	Asian
Coefficients:							
Age 18	-0.1102***	-0.0443	-0.1917***	-0.0865*	-0.1403	-0.1569	-0.2349
	(0.0421)	(0.0654)	(0.0509)	(0.0481)	(0.1110)	(0.1225)	(0.2016)
Age 17	0.0381**	0.0260	0.0526***	0.0309*	0.0936*	0.0222	0.0609
	(0.0162)	(0.0231)	(0.0202)	(0.0167)	(0.0544)	(0.0595)	(0.0860)
Age 16	-0.0090	-0.0078	-0.0104	-0.0068	-0.0268	-0.0172	-0.0253
	(0.0104)	(0.0114)	(0.0168)	(0.0114)	(0.0482)	(0.0469)	(0.0462)
Age 15	-0.0108	-0.0151	-0.0092	-0.0104	0.0525	-0.0386	-0.0042
	(0.0110)	(0.0108)	(0.0174)	(0.0114)	(0.0337)	(0.0398)	(0.0477)
F-test	22.84	22.77	22.83	20.10	29.10	16.21	25.95
Average marginal effects:							
Age 18	-0.0032***	-0.0010	-0.0070***	-0.0025*	-0.0063	-0.0033	-0.0105
Age 17	0.0011**	0.0006	0.0019***	0.0009*	0.0042*	0.0005	0.0027
Age 16	-0.0003	-0.0002	-0.0004	-0.0002	-0.0012	-0.0004	-0.0011
Age 15	-0.0003	-0.0003	-0.0003	-0.0003	0.0024	-0.0008	-0.0002
# of obs	1,373,608	784,689	588,919	1,101,135	85,070	94,601	56,343

Table 12: IV regression for relationship between majoring in STEM and H-1B population share when graduation.

% of H-1B population in:	All	Female	Male	White	Black	Hispanic	Asian
Coefficients:							
Age 22	-0.1636**	-0.1434*	-0.1952***	-0.1626**	0.0869	-0.1309	-0.7518**
	(0.0638)	(0.0861)	(0.0745)	(0.0682)	(0.1703)	(0.2260)	(0.3117)
Age 21	0.0445	0.0458	0.0468	0.0417	-0.0801	0.0649	0.3316**
	(0.0272)	(0.0348)	(0.0306)	(0.0275)	(0.0859)	(0.0924)	(0.1587)
Age 20	0.0115	0.0073	0.0161	0.0138	0.0349	-0.0552	0.0468
	(0.0110)	(0.0114)	(0.0144)	(0.0121)	(0.0407)	(0.0417)	(0.0434)
Age 19	-0.0248**	-0.0219*	-0.0289**	-0.0269**	-0.0152	0.0207	-0.0900
	(0.0108)	(0.0118)	(0.0135)	(0.0111)	(0.0429)	(0.0439)	(0.0551)
Age 18	0.0039	0.0130	-0.0067	0.0100	0.0358	-0.0872***	-0.0115
	(0.0088)	(0.0115)	(0.0124)	(0.0104)	(0.0358)	(0.0329)	(0.0395)
F-test	18.80	18.14	19.62	17.64	19.01	11.96	19.85
Average marginal effects:							
Age 22	-0.0021**	-0.0014*	-0.0032***	-0.0020**	0.0015	-0.0017	-0.0111**
Age 21	0.0006	0.0005	0.0008	0.0005	-0.0013	0.0008	0.0049**
Age 20	0.0001	0.0001	0.0003	0.0002	0.0006	-0.0007	0.0007
Age 19	-0.0003**	-0.0002*	-0.0005**	-0.0003**	-0.0003	0.0003	-0.0013
Age 18	0.0001	0.0001	-0.0001	0.0001	0.0006	-0.0011***	-0.0002
# of obs	1,627,098	925,293	701,805	1,312,411	102,196	108,507	61,990

3.7 Conclusion

This paper measures the possible impacts of H-1B visa reforms on US natives' college major choices in STEM related fields. In this study I have built a bridge between H-1B visa program related literatures and college major choices related literatures. Given the endogeneity problem of self-selection, I have constructed an instrumental variable of H-1B population share based on historical settlement pattern of foreign-born STEM workers. I find significant negative impacts of H-1B population shares on natives majoring in STEM fields, both when they enter the college and graduate from it. For students beginning their college education, a 10% increase in the H-1B population share decreases the probability of native students choosing STEM majors by 0.032%. If we take a further look at different gender and race subgroups, a 10% increase in the H-1B population share could decrease the likelihood of a male native student majoring in STEM related fields by 0.07%, and decrease that of a White native student by 0.025%. Given that the H-1B population share had been more than doubled from year 1992 to 2017, the probability of native students majoring in STEM fields when they enter the college would be 0.74 percentage points larger, if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Similarly, the likelihood of male native students choosing STEM majors would be 1.62 percentage points larger, and that of White native students would be 0.58 percentage points larger.

For students graduating from college, their reported majors are a joint choice of entering and shifting during college. Results indicate that a 10% increase in the H-1B population share decreases the probability of native graduates choosing STEM majors by 0.021%. When we take a further look at different subgroups, a 10% increase in the H-1B population share could decrease the likelihood of a female native graduate majoring in STEM related fields by 0.014%, decrease that of a male native graduate by 0.032%, and decrease that of a White native graduate by 0.02%. The native Asian subgroup suffer from the most dramatic crowd-out effect. A 10% increase in the H-1B population share could decrease the likelihood of an Asian native graduate majoring in STEM related fields by as large as 0.111%. Again, given that the H-1B population share had been more than doubled, the probability of native graduates majoring in STEM fields when they graduate would be 0.49 percentage points

larger, if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Similarly, the likelihood of female native graduates choosing STEM majors would be 0.32 percentage points larger, that of male native graduates would be 0.74 percentage points larger, and that of White native graduates would be 0.46 percentage points larger. For the native Asian subgroups, the probability would have been as large as 2.56 percentage points larger if the H-1B population shares had remained at their 1992 levels and all else had remained the same. Since foreign-born Asian account for a large proportion of H-1B visa holders, there might be an interesting "Asian crowd out Asian" story here.

Appendix A Additional Lemmas and Proofs for Chapter 2

The following lemma is a simplified version of Matrix Chernoff inequality, cited from [81, Eq. (5.1.5)].

Lemma 4. *Let $X_1, \cdots, X_m$ be a sequence of $d \times d$ i.i.d. positive semi-definite (PSD) matrices satisfying $\|X_k\|_{\text{op}} \leq L$ almost surely for all k. Then for any $t \in [0, 1)$,*

$$\Pr\left[\lambda_{\min}(\sum_j X_j) \leq t\mu_{\min}\right] \leq d \exp\left\{-\frac{(1-t)^2\mu_{\min}}{2L}\right\},$$

where $\lambda_{\min}(\cdot)$ is the smallest eigenvalue of a PSD matrix and $\mu_{\min} = \lambda_{\min}(\sum_j \mathbb{E}X_j)$.

The next lemma upper bounds the operator norm of an inverse covariance matrix of random vectors uniformly distributed on a unit sphere.

Lemma 5. *Let $u_1, u_2, \cdots, u_m$ be i.i.d. d-dimensional vectors uniformly distributed on the unit sphere $\{x \in \mathbb{R}^d : \|x\|_2 = 1\}$. Suppose also that $m \geq 8d\ln(d/\delta)$ for some $\delta \in (0, 1/2]$. Then with probability at least $1 - \delta$, $\|(\sum_j u_j u_j^\top)^{-1}\|_{\text{op}} \leq 2d/m$.*

Proof. Proof of Lemma 5. Denote $\Lambda := \mathbb{E}[u_1 u_1^\top] \in \mathbb{R}^{d\times d}$. Because the distribution of u_1 is spherical invariant, we immediately have $\Lambda_{jk} = 0$ for all $j \neq k$ and $\Lambda_{jj} \equiv \lambda$ for all $j = 1, 2, \cdots, d$. Additionally, since $\mathbb{E}[u_1^\top u_1] = 1$ we have that $\text{tr}(\Lambda) = d\lambda = 1$. This implies $\Lambda = \frac{1}{d} I_{d\times d}$.

Next, invoke Lemma 4 with $L = 1$, $\mu_{\min} = m/d$ and $t = 1/2$. We then have

$$\Pr\left[\lambda_{\min}(\sum_j u_j u_j^\top) \leq \frac{m}{2d}\right] \leq d \exp\left\{-\frac{m}{8d}\right\}.$$

Under the condition that $m \geq 8d\ln(d/\delta)$, the right-hand side of the above inequality is upper bounded by δ. Lemma 5 is thus proved. □

Appendix B Additional Data for Chapter 3

Table 13: STEM major classifications.

ACS Code	Major Name
1103	Animal Sciences
1104	Food Science
1105	Plant Science and Agronomy
1106	Soil Science
1301	Environmental Science
2001	Communication Technologies
21XX	Computer and Information Sciences
24XX	Engineering
25XX	Engineering Technologies
36XX	Biology and Life Sciences
37XX	Mathematics and Statistics
3801	Military Technologies
4002	Nutrition Sciences
4005	Mathematics and Computer Science
4006	Cognitive Science and Biopsychology
50XX	Physical Sciences
5102	Nuclear, Industrial Radiology, and Biological Technologies
5206	Social Psychology
6105	Medical Technologies Technicians
6108	Pharmacy, Pharmaceutical Sciences, and Administration
6202	Actuarial Science

Table 14: STEM occupation classifications.

Occupation	Occupation Codes			
	1980/90 census	**2000 census**	**2001-2009 ACS**	**2010-2017 ACS**
Actuaries	66	120	1200	1200
Aerospace engineers	44	132	1320	1320
Agricultural and food scientists	77	160	1600	1600
Airplane pilots and navigators	226	903	9030	9030
Atmospheric and space scientists	74	171	1710	1710
Biological scientists	78	161	1610	1610
Biological technicians	223	191	1910	1910
Chemical engineers	48	135	1350	1350
Chemical technicians	224	192	1920	1920
Chemists	73	172	1720	1720
Civil engineers	53	136	1360	1360
Clinical laboratory technologies and technicians	203	330	3300	3300
Computer software developers	229	102	1020	1020
Computer systems analysts and computer scientists	64	100	1000	1006
Dentists	85	301	3010	3010
Dietitians and nutritionists	97	303	3030	3030
Electrical engineers	55	141	1410	1410
Geologists	75	193	1930	1930
Industrial engineers	56	143	1430	1430
Management analysts	26	71	710	710
Mathematicians and mathematical scientists	68	124	1240	1240
Mechanical engineers	57	146	1460	1460
Medical scientists	83	165	1650	1650
Metallurgical and materials engineers, variously phrased	45	145	1450	1450
Not-elsewhere-classified engineers	59	153	1530	1530
Occupational therapists	99	315	3150	3150
Optometrists	87	304	3040	3040
Other health and therapy	89	326	3260	3260
Petroleum, mining, and geological engineers	47	152	1520	1520
Pharmacists	96	305	3050	3050
Physical scientists, n.e.c.	76	176	1760	1760
Physical therapists	103	316	3160	3160
Physicians	84	306	3060	3060
Physicians' assistants	106	311	3110	3110
Physicists and astronomers	69	170	1700	1700
Podiatrists	88	312	3120	3120
Psychologists	167	182	1820	1820
Sales engineers	258	493	4930	4930
Social scientists, n.e.c.	169	186	1860	1860
Speech therapists	104	323	3230	3230
Subject instructors (high school/college)	154	220	2200	2200
Therapists, n.e.c.	105	324	3240	3245
Veterinarians	86	325	3250	3250
Vocational and educational counselors	163	200	2000	2000

Bibliography

[1] A. Agarwal, O. Dekel, and L. Xiao. Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the Conference on Learning Theory (COLT)*, pages 28–40, 2010.

[2] A. Agarwal, H. Luo, B. Neyshabur, and R. E. Schapire. Corralling a band of bandit algorithms. *arXiv preprint arXiv:1612.06246*, 2016.

[3] J. G. Altonji, P. Arcidiacono, and A. Maurel. The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the Economics of Education*, volume 5, pages 305–396. Elsevier, 2016.

[4] C. Amuedo-Dorantes and D. Furtado. Settling for academia? h-1b visas and the career choices of international students in the united states. *Journal of Human Resources*, 54(2):401–429, 2019.

[5] P. Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004.

[6] P. Arcidiacono, E. M. Aucejo, and V. J. Hotz. University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review*, 106(3):525–62, 2016.

[7] P. Arcidiacono, E. M. Aucejo, and K. Spenner. What happens after enrollment? an analysis of the time path of racial differences in gpa and major choice. *IZA Journal of Labor Economics*, 1(1):5, 2012.

[8] P. Arcidiacono, V. J. Hotz, and S. Kang. Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166(1):3–16, 2012.

[9] M. Baird, M. Buchinsky, and V. Sovero. Decomposing the racial gap in stem major attrition: A course-level investigation. 2016.

[10] A. S. Bandeira, K. Scheinberg, and L. N. Vicente. Convergence of trust-region methods based on probabilistic models. *SIAM Journal on Optimization*, 24(3):1238–1264, 2014.

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

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

[13] O. Besbes, Y. Gur, and A. Zeevi. Non-stationary stochastic optimization. *Operations Research*, 63(5):1227–1244, 2015.

[14] R. Bhaskar, B. Arenas-Germosén, and C. Dick. Demographic analysis 2010: Sensitivity analysis of the foreign-born migration component. *US Census Bureau Population Division Working Paper*, 98:26–30, 2013.

[15] S. C. Billups, J. Larson, and P. Graf. Derivative-free optimization of expensive functions with computational error using weighted regression. *SIAM Journal on Optimization*, 23(1):27–53, 2013.

[16] J. R. Blum. Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics*, pages 737–744, 1954.

[17] S. Borenstein. Hubs and high fares: dominance and market power in the us airline industry. *The RAND Journal of Economics*, pages 344–365, 1989.

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

[19] S. Borenstein. An index of inter-city business travel for use in domestic airline competition analysis. *NBER working paper*, 2010.

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

[21] J. Bound, B. Hershbein, and B. T. Long. Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives*, 23(4):119–46, 2009.

[22] J. Bound and S. Turner. Coming to america: Where do international doctorate students study and how do us universities respond? In *American universities in a global market*, pages 101–127. University of Chicago Press, 2010.

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

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

[25] S. Bubeck, R. Munos, G. Stoltz, and C. Szepesvári. X-armed bandits. *Journal of Machine Learning Research*, 12(May):1655–1695, 2011.

[26] S. Bubeck, G. Stoltz, C. Szepesvári, and R. Munos. Online optimization in x-armed bandits. In *Advances in Neural Information Processing Systems*, pages 201–208, 2009.

[27] D. Card. Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64, 2001.

[28] D. Card and J. DiNardo. Do immigrant inflows lead to native outflows? *American Economic Review*, 90(2):360–367, 2000.

[29] D. Card and A. B. Krueger. Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of political Economy*, 100(1):1–40, 1992.

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

[31] R. Chen, M. Menickelly, and K. Scheinberg. Stochastic optimization using a trust-region method and random models. *Mathematical Programming*, 169(2):447–487, 2018.

[32] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to derivative-free optimization*, volume 8. SIAM, 2009.

[33] D. Costa and J. Rosenbaum. Temporary foreign workers by the numbers: New estimates by visa classification. *Economic Policy Institute, March*, 7, 2017.

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

[35] S. Dynarski. Building the stock of college-educated labor. *Journal of human resources*, 43(3):576–610, 2008.

[36] D. Escobari and P. Jindapon. Price discrimination through refund contracts in airlines. *International Journal of Industrial Organization*, 34:1–8, 2014.

[37] A. D. Flaxman, A. T. Kalai, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 385–394, 2005.

[38] GAO. Aviation competition: Restricting airline ticketing rules unlikely to help consumers. *Report to Congressional Committees, U.S.Government Accountability Office, Washington, DC.*, 2001.

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

[40] S. Ghadimi and G. Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

[41] S. Giaume and S. Guillou. Price discrimination and concentration in european airline markets. *Journal of Air Transport Management*, 10(5):305–310, 2004.

[42] P. Goldsmith-Pinkham, I. Sorkin, and H. Swift. Bartik instruments: What, when, why, and how. *American Economic Review*, 110(8):2586–2624, 2020.

[43] A. L. Griffith. Persistence of women and minorities in stem field majors: Is it the school that matters? *Economics of Education Review*, 29(6):911–922, 2010.

[44] J. Hunt. Which immigrants are most innovative and entrepreneurial? distinctions by entry visa. *Journal of Labor Economics*, 29(3):417–457, 2011.

[45] J. Hunt. The impact of immigration on the educational attainment of natives. *Journal of Human Resources*, 52(4):1060–1118, 2017.

[46] D. A. Jaeger, J. Ruist, and J. Stuhler. Shift-share instruments and the impact of immigration. Technical report, National Bureau of Economic Research, 2018.

[47] T. Kato and C. Sparber. Quotas and quality: The effect of h-1b visa restrictions on the pool of prospective undergraduate students from abroad. *Review of Economics and Statistics*, 95(1):109–126, 2013.

[48] W. R. Kerr and W. F. Lincoln. The supply side of innovation: H-1b visa reforms and us ethnic invention. *Journal of Labor Economics*, 28(3):473–508, 2010.

[49] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[50] J. F. Kirkegaard. Outsourcing and skill imports: Foreign high-skilled workers on h-1b and l-1 visas in the united states. Technical report, Working Paper, 2005.

[51] J. R. Koza. Genetic programming. 1997.

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

[53] R. M. Lewis and V. Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):1082–1099, 1999.

[54] D. Liu. A model of optimal consumer search and price discrimination in the airline industry, 2015.

[55] B. L. Lowell. H-1b temporary workers: Estimating the population. 2000.

[56] B. L. Lowell. Skilled temporary and permanent immigrants in the united states. *Population research and Policy review*, 20(1-2):33–58, 2001.

[57] B. L. Lowell and B. Christian. The characteristics of employers of h-1bs. *Institute for the Study of International Migration Working Paper*, 2000.

[58] M. Luppino and R. H. Sander. College major competitiveness and attrition from the sciences. *Available at SSRN 2167961*, 2013.

[59] T. Lykouris, V. Mirrokni, and R. Paes Leme. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122. ACM, 2018.

[60] MathWorks. Pattern search, accessed 3-october-2018, 2018.

[61] N. Matloff. On the need for reform of the h-1b non-immigrant work visa in computer-related occupations. *U. Mich. JL Reform*, 36:815, 2002.

[62] A. S. Nemirovsky and D. B. Yudin. *Problem complexity and method efficiency in optimization*. SIAM, 1983.

[63] P. M. Orrenius and M. Zavodny. Does immigration affect whether us natives major in science and engineering? *Journal of Labor Economics*, 33(S1):S79–S108, 2015.

[64] B. Ost. The role of peers and grades in determining major persistence in the sciences. *Economics of Education Review*, 29(6):923–934, 2010.

[65] G. Peri, K. Shih, and C. Sparber. Stem workers, h-1b visas, and productivity in us cities. *Journal of Labor Economics*, 33(S1):S225–S255, 2015.

[66] M. J. Powell. On trust region methods for unconstrained minimization without derivatives. *Mathematical Programming*, 97(3):605–623, 2003.

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

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

[69] K. Rask. Attrition in stem fields at a liberal arts college: The importance of grades and pre-collegiate preferences. *Economics of Education Review*, 29(6):892–900, 2010.

[70] M. Reksulak, W. Shughart, and G. Karahan. Vbarrier to entry: The po'litical economy of h'1b visasv. *V University of Mississippi Working Paper*, 2006.

[71] P. Robert S and R. Daniel L. Microeconomics, 2001.

[72] A. Saiz. Immigration and housing rents in american cities. *Journal of urban Economics*, 61(2):345–371, 2007.

[73] O. Shamir. An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, 18(52):1–11, 2017.

[74] O. Shy. *The economics of network industries*. Cambridge university press, 2001.

[75] C. L. Smith. The impact of low-skilled immigration on the youth labor market. *Journal of Labor Economics*, 30(1):55–89, 2012.

[76] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959, 2012.

[77] J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

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

[79] P. Surry. Hidden city ticket opportunities are more common than you think. *Hopper Research.*, 2005.

[80] V. Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, 1997.

[81] J. A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[82] P. J. Van Laarhoven and E. H. Aarts. Simulated annealing. In *Simulated annealing: Theory and applications*, pages 7–15. Springer, 1987.

[83] H. R. Varian. A model of sales. *The American economic review*, 70(4):651–659, 1980.

[84] Z. Wang and Y. Ye. Hidden-city ticketing: the cause and impact. *Transportation Science*, 50(1):288–305, 2016.

[85] D. Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4(2):65–85, 1994.

[86] F. Yu. Is anyone good enough for an h-1b visa? *The New York Times*, 2017.

[87] M. Zavodny and H. VThe. The h-1b program and its effects on information technology workers. *Economic Review-Federal Reserve Bank of Atlanta*, 88(3):33–44, 2003.