# Modeling Airlines' Route Selection Decisions Under Competition: A Discrete Games Based Model

Joseph D. Thekinen *
*School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 47906*

Kushal A. Moolchandani†
*School of Aeronautics and Astronautics, Purdue University, West Lafayette, Indiana, 47906*

Jitesh H. Panchal‡
*School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 47906*

Daniel A. DeLaurentis§
*School of Aeronautics and Astronautics, Purdue University, West Lafayette, Indiana, 47906*

**To analyze the effects of policies within the air transportation network, there is a need to model how policies affect the decisions made by airlines. Since airline decision making is based on proprietary information, such models need to rely on openly available data sources. In this paper, we use openly available data from the Bureau of Transportation Statistics to develop a predictive model of airline route selection decisions. The proposed model accounts for airline competition and models parameters such as operating cost, which can be influenced by policymakers. We illustrate the model using a dataset from two major airlines in US domestic Air Transportation Network. The dataset and the cost model are used for Bayesian estimation of model parameters, which are then used to predict the effects of cost and demand on the evolution of the network topology. The proposed model is found to be more accurate than competing models that do not consider the competition. From the estimates obtained on preference parameters, it is found that decreasing the operating cost and increasing the market demand increases the probability of operating service on the route for airlines, and the operating cost has a greater effect than market demand and route distance in the route selection decisions.**

*School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 47906
†Graduate Research Assistant, School of Aeronautics and Astronautics, Purdue University
‡Associate Professor, School of Mechanical Engineering, Purdue University, West Lafayette, Indiana, 47906
§Professor, School of Aeronautics and Astronautics, Purdue University,West Lafayette, Indiana, 47906

## Nomenclature

$A^1$ = Player 1

$A^2$ = Player 2

$r$ = route

$R$ = total number of routes in the ATN

$s_r^k$ = strategy of player $A^k$ on route $r$

$\vec{S}(r)$ = strategy profile $(s^1(r), s^2(r))$ in route $r$

$l^k$ = change in payoff of player $A^k$ due to the presence of its competitor

$U_{ij}^k(r)$ = utility achieved by player $A^k$ on operating route $r$ when strategy profile $(s^i, s^j)$ is followed

$V_{ij}^k(r)$ = observed component of utility $U_{ij}^k(r)$

$\vec{X}^k(r)$ = route characteristics of player $A^k$ in route $r$

$\vec{\beta}^k$ = preference parameter vector of player $A^k$ for route characteristics $\vec{X}^k$ in every route

$\vec{\beta}$ = $(\vec{\beta}^1, l^1, \vec{\beta}^2, l^2)$

$a^k(r)$ = airport presence of player $A^k$

$\vec{\alpha}^k$ = preference coefficient for airport presence of player $A^k$

$\vec{\theta}$ = $(\vec{\beta}, \vec{\alpha})$

# I. Introduction

Airline decisions on route selection depend on a large number of variables. The data on many of these variables are proprietary, and therefore not publicly available. By route selection, we refer to Origin-Destination (O-D) routes. This is not day of operations route planning, but establishing the O-D route plans in a schedule several months in advance. In the rest of the paper, by route selection we refer to O-D route selection. Several stakeholders, who do not have access to all the decision variables of the airlines, have an interest in understanding how airlines make these routse selection decisions. For example, regulatory bodies such as the Federal Aviation Administration (FAA) can use the entry decision model to direct the evolution of the Air Transportation Network (ATN) topology [1] towards improved connectivity, robustness, and resilience. Hence, there is a need to model such routing decisions using publicly available data.

In addition to being publicly available, it is important to base these decision models on factors that policymakers can influence. ATN evolves based on network decisions made by the airlines. If we construct a model that mimics network decisions of airlines by understanding their preferences towards factors that policymakers can influence, then such models can be used to provide useful guidance to regulatory bodies such as FAA to play an active role in channeling the network evolution towards targeted performance. For example, some of the factors such as operating cost can be directly influenced by the policymakers through incentives or imposing taxes. A decision model that estimates the preferences of airlines based on operating cost when making route decisions can be effectively used by policymakers in co-evolving the network.

Various approaches have been used in the past to model airline route selection decisions using publicly available data on route characteristics such as market demand, operating costs, and route length. These approaches were primarily based on approaches such as linear programming [2], integer programming [3], and machine learning [4]. All these approaches are focused on predicting the evolution of the ATN. Evans et al. [5] studied airline route selection behavior in response to congestion and delays at airports. Sha et al. [6] developed decision-based models to estimate the importance of decision variables that can be influenced by policymakers. These models are based on discrete choice, and use procedures such as maximum likelihood estimation to estimate the preferences of airlines towards route characteristics. The limitation of this model is that it does not account for the effect of competition among the airlines. In addition to route characteristics, potential competition from other airlines also influences these decisions. The profit that the airlines gain by operating a service on a route is affected by the presence of a competitor. The competitor may lead to a decrease in profit by reducing the market share or an increase in profit by reducing the overall operating cost. Based on the nature of interaction, the effect of the competitor on profit may be positive or negative. Doyme et al. [7] used an iterative procedure to model competition and simulate the decision variables of airfares, flight frequency, and choice of aircraft. However, the network structure was kept fixed and the prediction was on other decision variables of the airlines. A discrete games model developed by Bresnahan and Reiss [8] has been effectively used in the past to model the effect of competition on airline route selection decisions. These models study the effect of demographic factors such as city population of connecting airports or per capita income, the nature of destination such as vacation or commercial, and geographic factors such as distance. Policymakers do not have influence over these demographic and geographic factors. In summary, discrete choice-based models help to understand the preferences of airlines towards factors that the policymakers can influence but they do not account for competition. On the other hand, discrete games models [9] address competition but are limited in terms of usefulness to policymakers.

In this paper, we address these limitations by estimating preferences towards factors that the policymakers can influence, while including the effect of competition. Once the preferences are estimated, we study how the airline's entry decisions are affected by variations in these factors. In particular, we study how the strategic route decisions taken by airlines in the presence of competition from other airlines is affected by these variations. With the motivation of understanding how airlines make route selection decisions under competition, we answer the following research questions: (1) how can the preferences of airlines towards factors that the policy-makers can influence be estimated including the effect of competition? and (2) how do the variations in route characteristics affect the route decisions and the number of routes operated by the airlines?

To answer these questions, we build on the discrete games model by Ciliberto and Tamer [9] and estimate the preferences of airlines towards operating cost, market demand, and distance while making the route selection decisions. Operating cost, market demand, and distance are referred to as the route-attributes and airlines have preferences for these route-attributes. Our current model of preferences is a linear model with weights for route-attributes. The strategy of whether or not to operate a service on the route by each airline is based on the weightings for route-attributes, and presence of the competitor in the route. Estimating the preference parameters in the discrete games model is a major numerical challenge. Similar to Ciliberto and Tamer [9], we use a Bayesian-based approach to estimate these parameters. This involves obtaining the posterior distribution on the parameters. A posterior distribution is the probability distribution of an unknown quantity (which here are the preference parameters) conditional on the observed data (which here are the historic route selection decisions of the airlines). The Markov Chain Monte Carlo (MCMC) Metropolis-Hastings algorithm [10] method is used to sample the posterior. The presence of non-unique Nash equilibria makes the estimation of preference parameters challenging. Existing methods address this challenge by introducing a large number of latent parameters causing problems such as over-fitting. To tackle the problem of non-unique Nash equilibrium without over-fitting the model, we use a network level parameter called airport presence [11] in our MCMC method.

The remainder of this paper is organized as follows. Section II reviews related and relevant literature, Section III describes the theoretical model of airline competition and Section IV provides an overview of the numerical solution procedure used to estimate the parameters in the model. Section V describes the data used to run the route decision model. Section VI summarizes the estimates obtained for the preference parameters of the airlines. Section VII validates the model by evaluating the route prediction accuracy. In Section VIII , we conduct forward simulation using the predictive model to evaluate the effect of various policies. Concluding remarks are drawn on Section IX.

## II. Literature Review

Various approaches, either prescriptive or descriptive in nature, have been used in the past to model airline route selection decisions using publicly available data on route characteristics such as market demand, operating costs, route length (examples include Jaillet et al. [12], and Balakrishnan et al. [13], Lohatepanont and Barnhart [14]). Dantzig [15] and Kushige [2] used a prescriptive approach based on linear programming for fleet assignment on routes to maximize profits of airlines, although this approach breaks down when the number of alternative routes is large. Mixed integer programming was used by de Lamotte et al. [3] to study airline route schedule planning. Balakrishnan et al. [13] developed a Lagrangian-based solution to this mixed integer formulation of long-haul aircraft routing problems which selects candidate routes from a large set of possible alternatives. A category of descriptive approaches to model routing decisions of airlines is using network theory. Some of the examples of efforts using network theory are multiplier model by Song et al. [16], and machine learning techniques such as random forests, and support vector machine by Kotegawa [4] to model network evolution from historical data.

Most of the descriptive approaches mentioned above do not explicitly model the decisions made by airlines in response to factors that policy-makers can influence. Decision-based models have been developed to estimate the importance of decision variables that can be influenced by policy-makers. For example, Boguslaski et al. [17] studied the entry pattern of Southwest Airlines using demand, distance, and cost as route-attributes in making their decisions. Sha et al. [18] developed a decision-making model for a US-based airline using demand, operating costs, geographic distances between airports, and hub or non-hub nature of airports as the inputs variables. While Sha et al. [18] modeled demand as a continuously increasing input parameter, a later work by Moolchandani et al. [19] developed models of passengers' decision-making.

In addition to route characteristics, the potential competition from other airlines also influences these decisions. Berry [20] builds on the discrete games model developed by Bresnahan [8] to include competition between airlines using the logarithmic number of carriers operating service on a route as a measure of competition in the route. The main challenge while solving the aircraft routing problem using discrete games is the presence of non-unique Nash equilibria. Non-unique Nash equilibria makes it difficult to estimate the choice probabilities of strategy profile [21]. Ciliberto and Tamer [9] proposed a model based on Gibbs sampling procedure to overcome this challenge of non-uniqueness. A drawback of this model is that there are as many latent variables as there are routes in the network. The number of latent parameters grows in proportion to the square of the number of airports in the network leading to problems such as over-fitting and poor predictive accuracy.

In summary, there have been numerous studies on airline decision-making on route selection. To this literature we add consideration of competition and how it affects the airlines' choices. This model is an extension of the discrete choice model developed previously by the authors, see Ref. [6].

## III. Theoretical Model

This section describes the game theoretical approach that models the interaction between airlines while making route selection decisions. Following the work of Ciliberto et al. [9] the entry decisions made by airlines on each route are modeled as a discrete game with complete information played between two players. The payoff achieved by a player on operating a service on a route depends on route characteristics, preferences for route attributes and the presence of competitors in that route. Market factors are assumed to be known to the researcher. Here, the researcher is a stakeholder extraneous to the airlines who is trying to understand the route selection decision of the airlines and does not have access to all the private data that the airlines possess while making such decisions. The aim of the model is to estimate the airline preferences for route-attributes and the change in payoff due to the presence of the competitor, based on the available data of decisions made in the past. In this section, we present a theoretical model and an approach to estimate the parameters in the decision model. The section is structured as follows: Section III.A describes the game-theoretic model of the route-level interaction, Section III.B describes the forward model that characterizes the Nash equilibria strategies of the players, and Section IV is on the numerical procedure to obtain the posterior distribution on the target parameters based on the decisions made in the past.

### A. Game Theoretic Model of Route Interaction

Airline interaction at the route level is modeled as a discrete game of perfect information. Table 1 summarizes the payoff matrix of this game which describes the utility gained by each player for different strategy profiles. A strategy profile is a combination of strategies undertaken by the players. The strategy of a player $k$ is denoted by the symbol $s^k$, and on a route $r$, it is either to operate (denoted as ($s^k(r) = 1$)) or not to operate (denoted as ($s^k(r) = 0$)) service on that route.

| | | **(Player 2)** No route offered ($s_r^2 = 0$) | **(Player 2)** Route offered ($s_r^2 = 1$) |
|---|---|---|---|
| **(Player 1)** | No route offered ($s_r^1 = 0$) | $0, 0$ | $0, \vec{X}^2(r)\vec{\beta}^2 + \epsilon^2$ |
| **(Player 1)** | Route offered ($s_r^1 = 1$) | $\vec{X}^1(r)\vec{\beta}^1 + \epsilon^1, 0$ | $\vec{X}^1(r)\vec{\beta}^1 + l^1 + \epsilon^1, \vec{X}^2(r)\vec{\beta}^2 + l^2 + \epsilon^2$ |

**Table 1 Payoff matrix associated with the strategy profiles adopted by the players**

We assume that route decisions are independent, i.e., the decision on whether or not to operate service on a route is independent of the decisions made on other routes. However, it does depend on the connectivity and network-level influence of the route, which is accounted for in our model. From one time step to the next, airlines make decisions to add new routes or delete existing routes based on the utility attained. The utility attained by adding or deleting a route is affected by the decisions made on other routes. The network properties of a route such as the degree of the terminal airports, overall connectivity, and eigenvector centrality are affected by the decisions made on other routes. The network structure, in turn, alters the demand and cost of operating service on the route. However, from one year to the next, this effect on the change in utility is small as the fraction of routes added or deleted in a year is normally under 10%, as observed from the BTS T-100 dataset [22].

The utility attained by player $k$ on operating a service on the route is written as the sum of observed ($V_{ij}^k$) and unobserved ($\epsilon^k$) components, as shown in Eq. 1. The observed component accounts for all the factors observed by the researcher such as the market demand, route length, operating cost and presence of a competitor. The unobservable component is only from the perspective of the researcher. From the perspective of the players, this is a perfect information game.

$$U_{ij}^k(r) = V_{ij}^k(r) + \epsilon^k(r) \tag{1}$$

The observed component consists of two parts (Eq. 2): a) a route specific component that depends on the route characteristics ($\vec{X}^k(r)\vec{\beta}^k$), and b) a part that accounts for the effect of the presence of the competitor ($s^{-k}(r)l^k$).

$$V_{ij}^k(r) = \vec{X}^k(r)\vec{\beta}^k + s^{-k}(r)l^k \tag{2}$$

In the first component, ($\vec{X}^k$) denotes the route characteristics and ($\vec{\beta}^k$) denotes the preferences of player $k$ towards them. Following from Sha et al. [18], the main route characteristics contributing to airline decisions are market demand, non-stop distance, and operating cost. Therefore, these three route characteristics are considered in the model. In the second component, $s^{-k}(r)$ denotes the strategy adopted by a competitor of player $k$ on route $r$, and $l^k$ denotes the change in the utility of player $k$ in the presence of its competitor. The term $s^{-k}(r)l^k$, by formulation, vanishes for a route if the competitor decides not to operate service on that route. The parameter $l^k$ captures the difference in utility due to the presence of a competitor from being a monopolist operator on a route. This parameter is different for different airlines depending on the characteristics of the airlines and the market conditions. For example, full-service airlines and low-cost carriers respond differently in the presence of competition. $l^k$ is positive or negative depending on whether the presence of the competitor is beneficial or harmful. Sugawara and Omori [21] showed that for Japanese airline market JAL airlines favors those routes where ANA airlines have already been operating ($l^k > 0$ for JAL airlines) whereas entry of JAL airlines reduces the profit of ANA airlines ($l^k < 0$ for ANA airlines).

The strategy profiles of both the players and the utility associated with them are summarized as a payoff matrix in Table 1. In this matrix, each element has two values of payoff, where the first value corresponds to the utility of the first player, and the second value is the utility of the second player for the associated strategy profile. Without loss of generality, the utility of each player for not operating a route is assigned to be zero.

**B. Nash Equilibria of the Game**

The game described in Section III.A is played on each route. The rational strategy is to operate a service on the route if the payoffs are positive, and not operate otherwise. Following from Equations 1 and 2, the total utility attained by player $A^k$ on operating service on a route is obtained as $\vec{X}^k(r)\vec{\beta}^k + \; s^{-k}l^k + \epsilon^k$. The best response strategy of a player $A^k$ is to operate service on the route if the utility is positive as shown in Eq. 3. $s^k(r) = 1$ denotes that player $A^k$ operates a service on route $r$ and $s^k(r) = 0$ denotes that player $A^k$ does not operate a service on route $r$.

$$s^k(r) = \begin{cases} 0 & \text{if} \;\; \vec{X}^k(r)\vec{\beta}^k + s^{-k}l^k + \epsilon^k \leq 0 \\ 1 & \text{if} \;\; \vec{X}^k(r)\vec{\beta}^k + s^{-k}l^k + \epsilon^k > 0 \end{cases} \tag{3}$$

There are four possible pure Nash equilibria strategies corresponding to strategy profiles $\vec{S}(r) = (s^1(r), s^2(r)) = (0,0), (1,0), (0,1), (1,1)$. For a given payoff matrix, a unique Nash equilibrium strategy may or may not exist depending on the payoffs associated with the strategy profiles. Assuming that the presence of the competitor negatively affects the utility of operating service on the route for both the players $l^1 < 0, l^2 < 0$, the Nash Equilibrium of the game in each route is expressed as a function of unobserved variables, $\epsilon^1, \epsilon^2$ as shown in Figure 1. This is a reasonable assumption as the entry of an airline in a route decreases the profit of the other airlines operating in that route in the US domestic ATN [21]. The regions of Nash-equilibria are different if the presence of a competitor increases the utility attained for any one player. A detailed analysis of all such cases is provided by Bresnahan and Reiss [8].

Figure 1 is a plot of Nash-equilibria of the game described in Table 1 with unobserved variables as axes ($\epsilon^1$ as the horizontal axis and $\epsilon^2$ as the vertical axis). Regions A, B, C, and D have unique Nash equilibria corresponding to equilibria $\vec{S}(r) = (s^1(r), s^2(r)) = (0,0), (1,0), (0,1), (1,1)$ respectively. Region E has multiple Nash-equilibria.

The presence of regions with multiple Nash-equilibria makes the estimation of model parameters harder. This is because if we assign distributional assumptions on the unobserved variables $(\epsilon^1, \epsilon^2)$ the choice probabilities on Nash equilibria are not well-defined in the regions of multiple equilibria. We propose an approach based on a network parameter called airport presence [11] to overcome this problem of regions of multiple Nash equilibria. The airport presence of an airline at a given airport is the ratio of the number of airports to which an airline operates a direct flight from the given airport to the total number of airports to which all airlines operate a direct flight from the given airport. We assume that the probability of a player entering the route conditional on the route falling in region E is a function of the airport presence of the player in the route. Several authors (e.g., Borenstein [23], and Levinine [24]) have argued that the level of operations in an airport has a significant impact on the competitive position on the route operated from or to the airport. Berry [11] investigated the effect of airport presence on the oligopoly product differentiation. Our

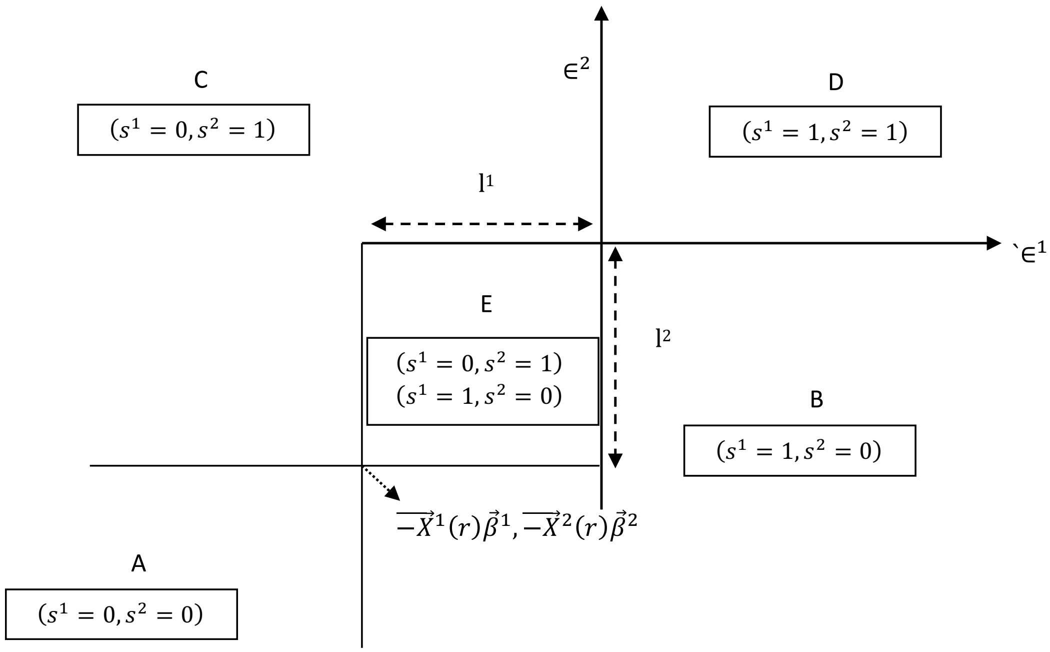


**Fig. 1 Nash Equilibria regions as functions of unobserved variables assuming** $l^1 < 0, l^2 < 0$

examination of historic decisions indicates that the probability of a player entering is a non-linear function of airport presence. We observe from the past data [22] that if the airport presence of a route was more than a threshold (0.3 for the year shown in the figure) then the probability of retaining that route was quite high as observed from data. As an example, Figure 2 shows the entry decision of United Airlines (UA) as a function of its airport presence in the year 2012-13. All routes in a network formed by the top 132 US domestic airports [25] was considered in this plot. A similar logistic functional relation between entry decision of UA and airport presence was observed for other years as well. Therefore, in region E, a logistic functional relation was assumed between the probability of playing equilibrium $(1, 0)$ and airport presence, as shown in Equation 4.

$$p_r = (\alpha_1 + \alpha_2 a^1(r))^{-1} \tag{4}$$

This probability of equilibrium $(1, 0)$ being played in route $r$ is denoted by $p_r$ and airport presence of UA is denoted by $a^1(r)$. In Equation 4, $\alpha_1, \alpha_2$ are parameters that will be estimated using the procedure described in Section IV.

## IV. Numerical Procedure to Estimate the Parameters in the Theoretical Model

To estimate the parameters of the decision model developed in Section III, a Bayesian approach is used with priors obtained using Discrete Choice Analysis (DCA) and likelihood obtained by data from historic decisions made by

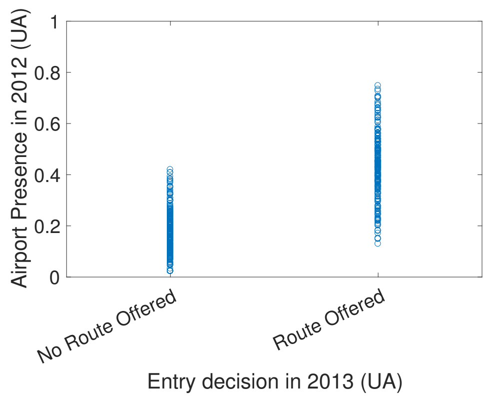


**Fig. 2 Entry decision of United Airlines (UA) on each route as a function of airport presence for that route in the year 2013.**

the airlines. The posterior distributions of parameters, so obtained, are sampled using the standard MCMC method based MH algorithm [10]. The overall procedure is summarized in Figure 3. A Bayesian-based numerical approach is used to estimate the preference parameters of the discrete games model. The priors are obtained using DCA ignoring competition which is described in Section IV.A. The data on historic decisions made by individual airlines on different routes and the associated route-attributes is used to obtain the likelihood functions. The likelihood formulation is described in Section IV.B. A posterior on the preference parameters of the airlines is formulated using the prior and posterior using a Bayesian approach, which is discussed in Section IV.C. Section IV.D briefs the MCMC-MH algorithm that is used to sample the posterior and estimate the preference parameters.

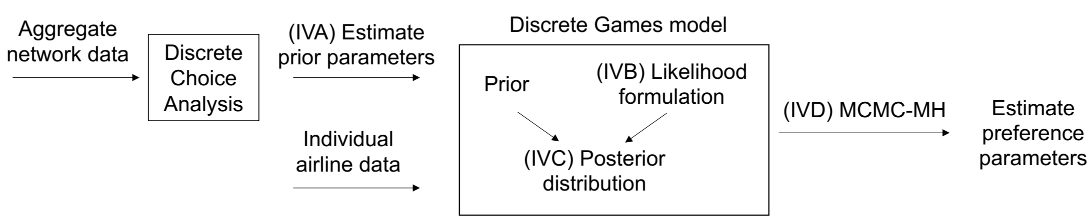


**Fig. 3 Overview of procedure to estimate airline preferences using network data**

### A. Priors using Discrete Choice Analysis

Sha et al. [6] developed a model based on discrete choice random-utility theory to estimate decision-making preferences of airlines. In DCA, the utility ($U_i$) for alternative $i$ consists of a component that is observed by the researcher ($V_i$) and unobserved component $\epsilon_i$ which is the uncertainty, as shown in Eq. 5.

$$U_i = V_i + \epsilon_i \tag{5}$$

The observed component is deterministically estimated from the researcher's point of view through different techniques such as survey data and expertise. The unobserved component captures the uncertainty due to unobserved attributes, measurement errors, etc. For example, the airline route decision criteria not included in the model are captured in this term. A linear form of the observed component was used by Sha et al. [6] as shown in Eq. 6.

$$V_i = \vec{x}^T \vec{\beta}_i \tag{6}$$

where $\vec{x} = (x_1, ..., x_n)^T$ is a set of $n$ route-attributes for utility, and $\beta_i = (\beta_{i1}, ..., \beta_{ik})^T$ is a set of weights that quantify the preferences of decision maker $i$. Under random-utility maximization assumption, the decision maker prefers alternative $i$ over $j$ if $U_i \geq U_j$. The probability of a decision maker choosing alternative $i$ is obtained as a function of the cumulative distribution of $(\epsilon_j - \epsilon_i)$ as shown in Eq. 7 [26].

$$P_i = P(U_i \geq U_j) = P(V_i - V_j \geq \epsilon_j - \epsilon_i) \tag{7}$$

Sha et al. used a multinomial logit model [27] that assumes $\epsilon_i$ to be independent and identically distributed following a Gumbel distribution. The route-attributes were market demand, direct operating cost (DOC), distance and whether a route connects hub airports or not. The preferences for the route-attributes were then estimated using the discrete choice model.

In our model, we use the estimates of the preferences obtained by running the discrete choice model as parameters of the prior distribution. DCA gives point estimates for preference parameters of the airlines. For preference parameters of each player towards the route-attributes, a normal prior is assigned with mean obtained using DCA. For the parameter $l^k$, a normal prior is assigned truncated in the region $(-\infty, 0)$ since we assume that the presence of the competitor decreases the utility for both the players. $\alpha_1, \alpha_2$ are given a normal prior. The parameters for prior on $\vec{\alpha}$ are obtained by fitting a logistic curve between airport presence and entry decision of Player $A^1$ in all routes where $(0, 1), (1, 0)$ equilibria were played.

### B. Likelihood

After assigning the priors, the next step is to obtain likelihood function of the model parameters from the observed data. Assuming that the unobserved variables, $\epsilon^1$ and $\epsilon^2$, are independent and follow the standard normal distribution, the likelihood of each of the strategy profile regions in Fig. 1 is expressed as a function of the vector of route characteristics $\vec{X}^k(r)$ and the preferences $\vec{\beta}^k$ towards them. Equations 8–12 are the likelihood values of falling in each of the five regions [21]. In these equations, $\phi$ denotes cumulative density of standard normal distribution and $L^i$ denotes the likelihood of falling into region $i$ in Figure 1.

$$L_A(\vec{\beta}, \vec{X}_r) = \phi(-\vec{X}_r^1\vec{\beta}^1)\phi(-\vec{X}_r^2\vec{\beta}^2) \tag{8}$$

$$L_B(\vec{\beta}, \vec{X}_r) = \{\phi(\vec{X}_r^1\vec{\beta}^1) - \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\}\phi(-\vec{X}_r^2\vec{\beta}^2) + \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\phi(-\vec{X}_r^2\vec{\beta}^2 - l^2) \tag{9}$$

$$L_C(\vec{\beta}, \vec{X}_r) = \{\phi(\vec{X}_r^1\vec{\beta}^1) - \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\}\phi(\vec{X}_r^2\vec{\beta}^2 + l^2) + \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\phi(\vec{X}_r^2\vec{\beta}^2) \tag{10}$$

$$L_D(\vec{\beta}, \vec{X}_r) = \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\phi(\vec{X}_r^2\vec{\beta}^2 + l^2) \tag{11}$$

$$L_E(\vec{\beta}, \vec{X}_r) = \{\phi(\vec{X}_r^1\vec{\beta}^1) - \phi(\vec{X}_r^1\vec{\beta}^1 + l^1)\}\{\phi(\vec{X}_r^2\vec{\beta}^2) - \phi(\vec{X}_r^2\vec{\beta}^2 + l^2)\} \tag{12}$$

The joint likelihood function assuming that routes are independent is given by Equation 13.

$$L(\vec{X}, \vec{S}|\vec{\beta}, \vec{\alpha}, l^1, l^2) = \prod_{r=1}^{R}\left[\left\{(L_A)^{I[\vec{S}(r)=1]}(L_B + \lambda_r L_E)^{I[\vec{S}(r)=2]}(L_C + L_E - \lambda_r L_E)^{I[\vec{S}(r)=3]}\right.\right.$$
$$\left.\left.(L_D)^{I[\vec{S}(r)=4]}\right\}\left\{\frac{1}{1 + e^{\alpha_1 + \alpha_2 a^k}}\right\}^{I[\vec{S}(r)=2,3]}\right] \tag{13}$$

In this equation, $R$ is the total number of routes in the network and $L_i$ are defined in Eqs. 8–12. $\lambda_r$ is a Bernoulli variable with airport presence as the parameter (Equation 14).

$$\lambda_r \sim \ Bernoulli(p_r) \tag{14}$$

$\lambda_r = 1$ when $(1, 0)$ equilibrium played and $\lambda_r = 0$ when $(0, 1)$ is played.

**C. Posterior**

The joint posterior density function for the parameters $\vec{\alpha}$ and $\vec{\beta}$ is proportional to the product of prior and likelihood as shown in Eq. 15.

$$Y(\vec{\beta}, \vec{\alpha}, l^1, l^2|\vec{X}, \vec{S}) \propto \pi(\vec{\alpha})\pi(\vec{\beta})\pi(l^1)\pi(l^2)L(\vec{X}, \vec{S}|\vec{\beta}, \vec{\alpha}, l^1, l^2) \tag{15}$$

where $\pi(\vec{\beta})$, $\pi(\vec{\alpha})$, and $\pi(\vec{l^k})$ denote the prior distribution on preference parameters on route-attributes, airport presence, and interaction factor respectively, and $L$ is the joint likelihood function.

**D. Sampling posterior distribution**

The posteriors on the parameters $\vec{\beta}$, $l^1$, and $l^2$ are sampled using the standard MCMC method based MH algorithm [10]. The algorithm proceeds through a random walk through the sample space. It proceeds such that the distribution of the next sample depends on the current sample. The next sample is accepted or rejected with some probability specified by something known as proposal distribution. A jump function is used to define the proposal distribution. A distribution similar to the prior with the same mean and standard deviation of 25 times lower than the prior was used as the jump function in the MH algorithm. The more probable the next sample, higher the probability of it being accepted. This iterative procedure generates a sequence of values and the probability distribution function of the sequence of values will asymptotically approach the posterior distribution upon running sufficient number of times.

## V. Data for Empirical Study

Among the many possible abstractions of the US Air Transportation System (ATS), we consider a network formed by the top 132 US domestic airports sorted in the decreasing order of passenger demand. These are the largest commercial airports as listed by FAA and handle more than 95 percent of all the enplanements [25]. The number of plausible routes grows in proportion to the square of the number of airports. Thus, our network has a total of 8646 possible routes among the airports considered. We use BTS data from the T-100 data bank [22] for data on airline operations and schedule P-5.2 [28] for data on airline economics. We use both the segment and market data from the T-100 data bank, which gives us information such as the number of enplaned passengers, departures performed, and origin and destination airports. The scheduel P-5.2 contains detailed quarterly aircraft operating expenses for large certificated US air carriers and provides us with information on the airlines' total direct operating expenses.

The raw data obtained from the above source requires some cleaning before it can be used for analysis. Hence, we applied a set of filters. First, for an airline, we considered only those routes to exist where the airline had at least eight scheduled departures in any two consecutive months of a year with a nonzero passenger demand in the domestic category. This condition was chosen because of the following reasons: a) freight aircraft were eliminated considering only those with nonzero passenger demand, b) seasonal effect was removed by considering routes that are operated across all seasons in a year by considering any two consecutive months, c) there should be at least one scheduled departure a week for the minimum two month period that the route was operational so that those routes that were operated for special purposes and which were not part of the original route are eliminated from the analysis, and d) only domestic operations were considered.

We considered the competition between two major legacy carriers in the US domestic network – United and Delta – and removed data for the other airlines. These airlines will be referred to as player $A^1$ (UA) and $A^2$ (DL) respectively in rest of the paper. Among the 8646 possible routes, UA operated on 468 routes and DL operated on 328 routes for at least one year from the years 2005-14 based on BTS data from T-100 data bank [22]. During the same period they simultaneously operated service on 57 routes competing for the market share. We assume that both these airlines compete with each other and each player makes a decision to enter a route based on the route characteristics and expectation on the strategy of its competitor.

Route passenger demand, DOC of the route, and distance were the route characteristics chosen in the model. BTS T-100 data bank [22] was used to obtain route passenger demand. There are two kinds of demand: market demand and segment demand. Market demand includes only those passengers that had the terminal airports of a route as both origin and destination. Segment demand on a route counts all the passengers that used that route as a leg in the trip. We used market demand as a measure of route passenger demand. We calculate the DOC over a route for an airline weighted by the number of operations. In certain routes the cost values are missing from the dataset in certain years. We use an airline cost index based on data from the Statistical Abstract of United States [29] to estimate them. We assume passenger demand and DOC to be symmetric by calculating the average of bidirectional value for every airport pair. For the remaining of the manuscript the DOC will be referred to as cost.

## VI. Estimation Results of the Empirical Study

This section shows the results of Bayesian estimation on preference parameters of the discrete games model.

The results from the DCA were used to obtain the parameters for the prior distribution. DCA provides separate preference parameter values for route addition and route deletion. A weighted average of this (based on the fraction of operating and non-operating routes) were used to obtain the prior parameter. The standard deviation was set high so that the prior remains flat and does not bias the posterior. Thus the resulting posterior distribution is inferred primarily from the real data. The parameters of the prior distribution for the three route-attributes are tabulated in Table 2. Prior Mean (add) are the results obtained on preference parameters by running the discrete choice model for route addition. Prior Mean (del) are the same results for route deletion. The DCA does not distinguish between airlines and therefore the same parameter values were used in prior for both the airlines.

**Table 2 Statistics of the posteriors of decision parameters (after burn-in period).**

| Parameter | Prior mean | Prior stdev |
|---|---|---|
| Demand | 0.055 | 10 |
| Cost | 0.469 | 10 |
| Distance | -0.044 | 10 |

A total of 1 million samples were used in the Metropolis-Hastings algorithm to estimate the posterior distribution of the preference parameters. For the proposed distribution a standard deviation 5000 times smaller than the prior was used. This resulted in a reasonable acceptance rate of 33.4% for the posterior samples when Metropolis-Hastings MCMC was implemented. The raw MCMC posterior samples for the preference parameter towards the route-attributes of both the airlines are shown in Figure 4.

$200,000$ samples were taken as the burn-in period. In MCMC-MH algorithm implementation, a common phenomenon is presence of correlated successive sample. Samples that are correlated with one another do not convey extra information and needs to be removed. To achieve this after removing the burn-in period, autocorrelation plots were made for the remaining samples. Figure 5 shows the average autocorrelation for all the posterior samples as a function of samples that are $lag$ spaced apart. For smaller values of $lag$ the average autocorrelation is high as adjacent samples

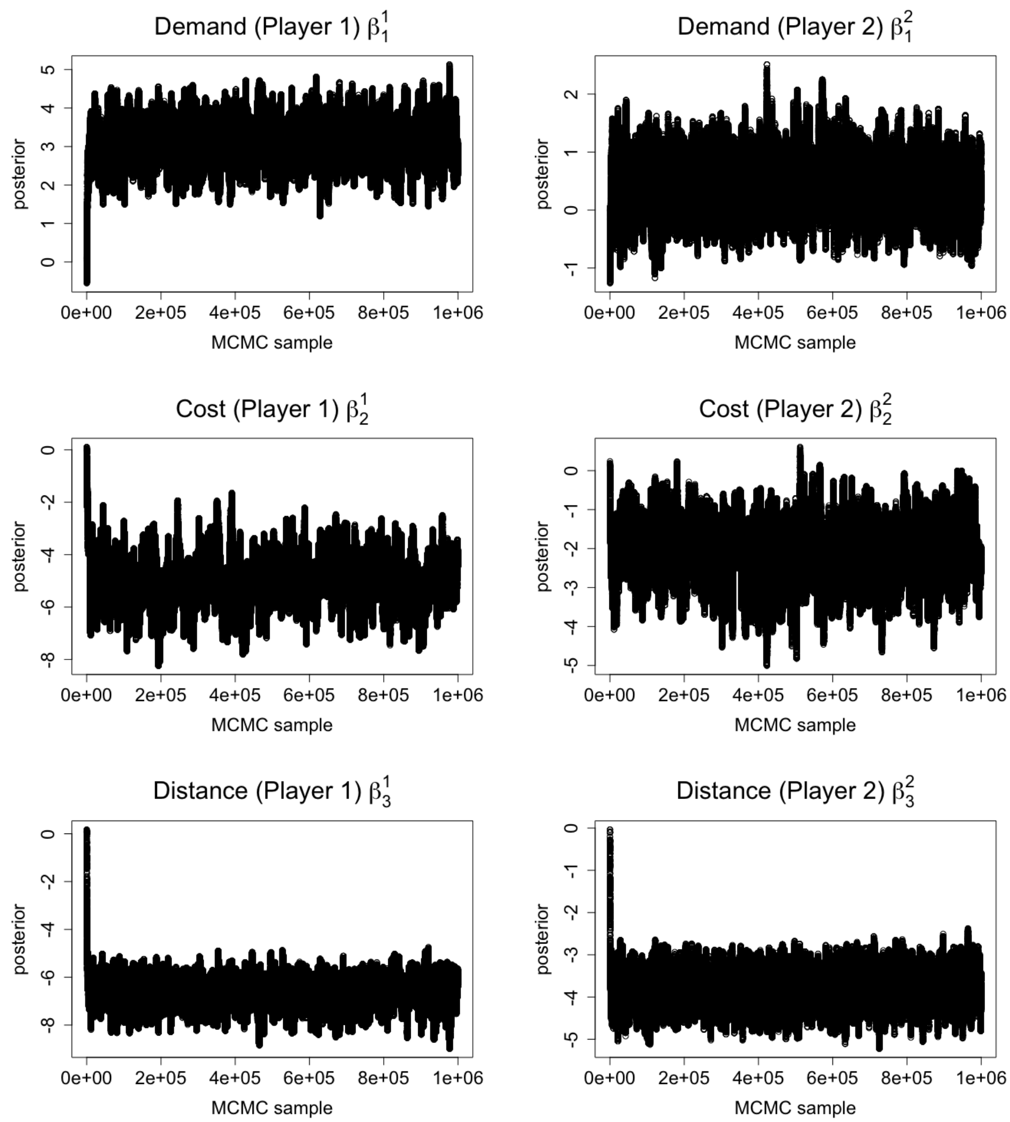


**Fig. 4 Raw MCMC posterior samples of the preference parameters towards the route route-attributes.**

tend to be highly correlated. $lag$ should be chosen such that the autocorrelation between samples is low. Figure 5 shows that when $lag$ is $10,000$ the autocorrelation values are low for all the posterior samples. Therefore, one in every $10,000$ sample is picked to obtain the posterior distribution.

Table 3 compares the statistics of the posterior on preference parameters with estimates using DCA for the same year. DCA provides two sets of parameters, one for non-operating routes for route addition and the other for operating route for route deletion. The posterior samples on preference parameters quantify the preferences of the airline towards route-attributes such as cost (unit: cent/nautical mile/seat), demand (unit: 1000 passengers), and distance (unit: 1000 nautical miles). For the interaction parameter the prior parameters correspond to the intercept of DCA result. For the coefficient of airport presence the prior parameters were obtained by curve-fitting the parameters $\alpha_1$ and $\alpha_2$ (Equation 4)

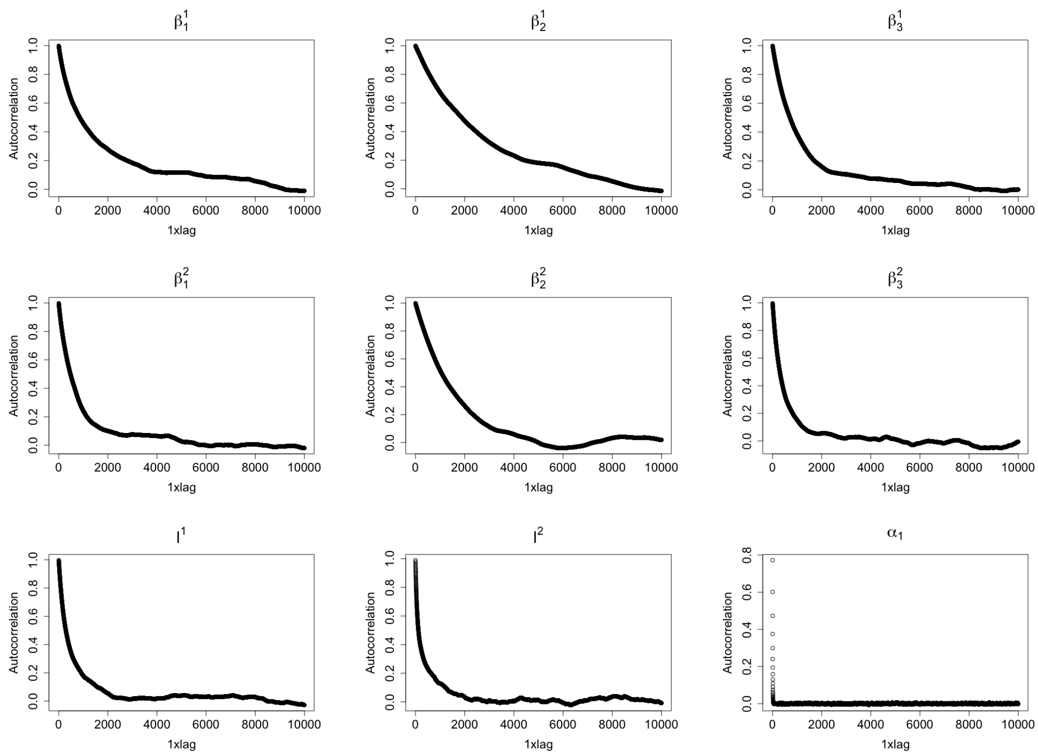


**Fig. 5 Plot showing average autocorrelation between samples that are *lag* spaced apart for each all the parameters.**

on the data for the previous year.

**Table 3 Statistics of the filtered posterior samples.**

| Parameter | DCA (add) | DCA (del) | Posterior mean (UA, $k = 1$) | Posterior mean (DL, $k = 2$) | Posterior std dev (UA) | Posterior std dev (DL) |
|---|---|---|---|---|---|---|
| Demand ($\beta_2^k$) | 0.087 | -0.23 | 3.11 | 0.47 | 0.49 | 0.38 |
| Cost ($\beta_1^k$) | 0.74 | -1.97 | -5.01 | -2.17 | 0.93 | 0.65 |
| Distance ($\beta_3^k$) | -0.04 | -0.17 | -6.66 | -3.84 | 0.44 | 0.37 |
| Interaction ($l^k$) | -1.91 | 1.49 | -1.22 | -0.17 | 0.35 | 0.18 |

Table 3 shows that both the players under competition have a positive coefficient for demand and negative for the cost and distance. This implies that, according to the model, higher the demand that a route brings in and lower the cost involved in adding the route, higher is the utility of adding that route. Thus increasing cost negatively affects the probability of the route getting added. Furthermore, the magnitude of preference parameter for cost was one order of magnitude higher than the parameter for demand which means that cost had a higher significance in the route selection decision. This is encouraging from policy designer standpoint as cost is the variable that they can directly control. Demand is governed more by market factors. The interaction parameters, $l^1$ and $l^2$ denotes the decline the effect of presence of a competitor for UA and DL respectively. As observed from Table 3, interaction parameter for both UA and DL are negative which means that the presence of the competitor adversely affects the utility of both the airlines. Furthermore, UA had a higher magnitude of interaction parameter than DL. This implies that UA had a larger decline in the payoff from the presence of the competitor as compared to DL for the range of years considered.

## VII. Route prediction accuracy

To evaluate the discrete games model, route prediction accuracy was measured and compared with the discrete choice model. Route prediction accuracy of an airline is defined as the fraction of routes where the model accurately predicts the entry decision of the airline. To obtain this accuracy a route-by-route comparison of the model prediction and the actual decision was made. For the period 2006-07, the model was trained using decisions made by both the airlines on 500 routes (of the 8646) in the year 2006. These 500 routes were a random sample consisting of all the four Nash equilibria. The trained model was then used to predict the accuracy in the year 2007. A comparison of the prediction accuracy obtained using the current model with that of discrete choice is tabulated in Table 4. The overall accuracy of prediction by DCA from years 2006-13 is 90.6% and the average accuracy for the current model is 89.2% for UA and 81.8% for DL for prediction during the years 2006 to 2013.

**Table 4 Comparison of overall prediction accuracy of DCA and current model.**

| Period | Discrete choice | UA (current) | Delta (current) |
|---|---|---|---|
| 2006-07 | 90.4% | 89.1% | 85.9% |
| 2007-08 | 89.3% | 89.3% | 85.8% |
| 2008-09 | 89.6% | 90.3% | 86.7% |
| 2009-10 | 91.8% | 89.8% | 78.0% |
| 2010-11 | 90.1% | 90.4% | 79.2% |
| 2011-12 | 90.1% | 88.0% | 79.3% |
| 2012-13 | 90.5% | 87.2% | 77.6% |

Though the overall accuracy of discrete choice is better than the current model, in routes where an addition or deletion decision was made (we call this dynamic accuracy) the current model performs better than the discrete choice model. The results for the dynamic accuracy are tabulated in Table 5. This is because discrete choice prediction bases itself on the current status of the route. It relies on two separate models – one for route addition and the other for route deletion. Implicitly, the information of the current status of the route increases the prediction accuracy achieved using discrete choice model. From one period to the next, addition or deletion happens only in less than 10% of the routes and the discrete choice model predicts that the route status is unchanged in a significant majority of the routes. For example, in the period 2006-07 among the 2591 existing routes, only 53 routes were selected for deletion and only 107 new routes were added to the network. Even a model that takes information about route status in the previous year and predicts no changes to happen would yield a 94% accuracy. But the real use of the model is to understand how airlines make routing decisions. Therefore the merit of the model is gauged by comparing the accuracy only on those routes where route status changed from one period to the next. We call this dynamic accuracy of the model. The model that predicts no changes in the route status will have a dynamic accuracy of 0%. The discrete choice model had an average dynamic accuracy of 50.3%, and the current discrete games model has an average dynamic accuracy of 68.9% for UA and 69.2% for DL during the years 2006 to 2013. The accuracy achieved in individual years is listed in Table 5. This increased accuracy justifies that including the effect of competition helps better understand how airlines make addition and deletion decisions. The increased overall accuracy of DCA is due to the information of the current route structure. In future work, it would be interesting to study whether including other route-attributes that provide information about the current route structure improve the overall accuracy of the model.

**Table 5 Comparison of dynamic prediction accuracy of DCA and current model.**

| Period | Discrete choice | UA (current) | Delta (current) |
|---|---|---|---|
| 2006-07 | 50.5 % | 70.9 % | 74.9 % |
| 2007-08 | 40.0 % | 65.7 % | 74.3 % |
| 2008-09 | 55.3 % | 75.3 % | 76.5 % |
| 2009-10 | 59.1 % | 77.9 % | 75.6 % |
| 2010-11 | 43.0 % | 62.3 % | 59.4 % |
| 2011-12 | 47.4 % | 65.6 % | 60.4% |
| 2012-13 | 56.7 % | 64.9 % | 63.6 % |

Among the 8646 routes considered in our network, no O-D route existed for more than 80% of the routes. A trivial model that says no O-D route will be added can achieve accuracy higher than 80%. Therefore, the prediction accuracy metric alone is not sufficient to assess the performance of the model. Receiver Operating Characteristic (ROC) curves are used to judge the discrimination ability of the predictor [30]. The most quantitative index describing an ROC curve is the area under the curve and this area is termed as Area Under Curve (AUC) [30]. ROC curves and AUC metric are popular measures to evaluate the model because they convey information about discrimination ability of the model. A trivial model that says no route gets added generates more than 80% prediction accuracy, but AUC metric will evaluate such a model poorly as the model has poor discrimination ability. ROC curves and AUC metric evaluates the discrimination ability of the predictive model. Figure 6 shows the ROC curves for prediction of route entry decision made by Player 1 (UA) and Player 2 (DL) separately. The predictive model for Player 1 had an AUC of 0.77, whereas Player 2 our model attained an AUC of 0.70.

## VIII. Policy experimentation discussion

Based on the estimates for airlines' preferences of route-attributes such as demand, cost, and distance the evolutionary behavior of the ATN is studied in this section. In particular, we focus on the effect of the route-attributes on the

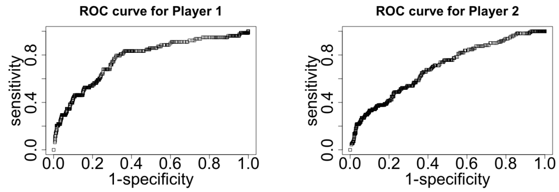


**Fig. 6 ROC curve for prediction of route entry decision.**

Nash-equilibrium.

### A. Forward simulation

In Section VII we validated the prediction accuracy of the model based on the estimated preferences towards route-attributes. Since the model is able to predict the airline route decisions using only the route-attributes with high accuracy we hypothesize that the model should capture the decision-making behavior of the airlines on varying the route-attributes. In this section we show the effect of varying the route-attributes on the airline routing decisions.

As the non-stop route distance cannot be varied, the two route-attributes that were considered in our analysis were market demand and cost. While varying the route-attributes, it was performed at a network level, i.e., all the routes are simultaneously varied and network level aggregated effects are studied. While varying the cost the change was uniform for every route whereas for market demand the change for a route was in proportion to the existing demand in that route. This is because for passenger demand variations between period are usually in proportion to the congestion. For example, the busier an airline route, the more is the absolute value of the demand fluctuation. On the other hand, for cost if airport charges or landing fees is increased then it is absolute for all routes and not in proportion to the existing cost for a route. However, this assumption of relative or absolute does not fundamentally change the observation and conclusions drawn from varying the route-attributes. The assumption of distinguishing proportional from absolute was made only to mimic the variations in practice while performing the forward simulation.

#### *1. Varying cost*

Based on the decisions made by the two airlines (whether to operate or not to operate), there are four possible outcomes or Nash equilibria strategies in each route. These are $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$; where the first entry denotes the strategy of Player 1 and second entry denotes the strategy of Player 2. Figure 8 shows the number of routes where each of the four Nash equilibria strategies were played as a function of cost incurred by Player 2. The cost is increased uniformly for all the routes. In the figure, ‘std dev’ indicates the standard deviation in the distribution of cost across all routes in the network the network. Along the horizontal axis is the number of standard deviations by which the costs were increased for Player 1.

Before discussing the significance of these results, we explain the Nash equilibrium behavior with varying cost

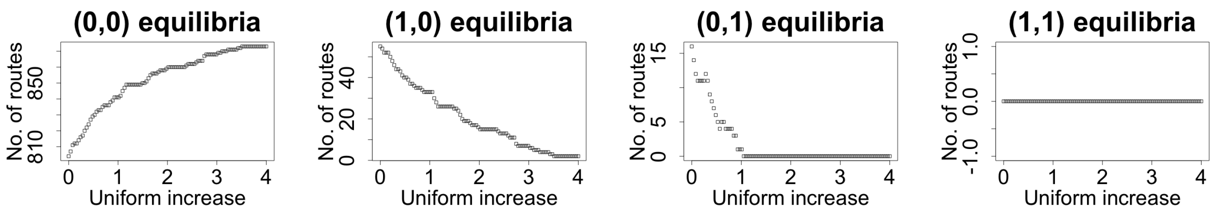


**Fig. 7 Comparing the number of routes in each Nash-equilibrium was predicted by increasing the cost of Player 1 (UA).**

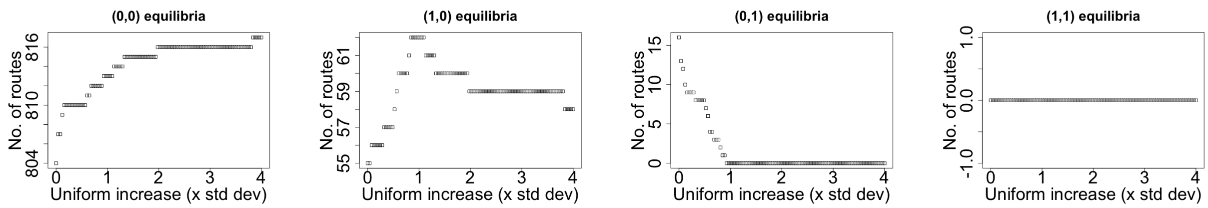


**Fig. 8 Comparing the number of routes in each Nash-equilibrium was predicted by increasing the cost of Player 2 (DL).**

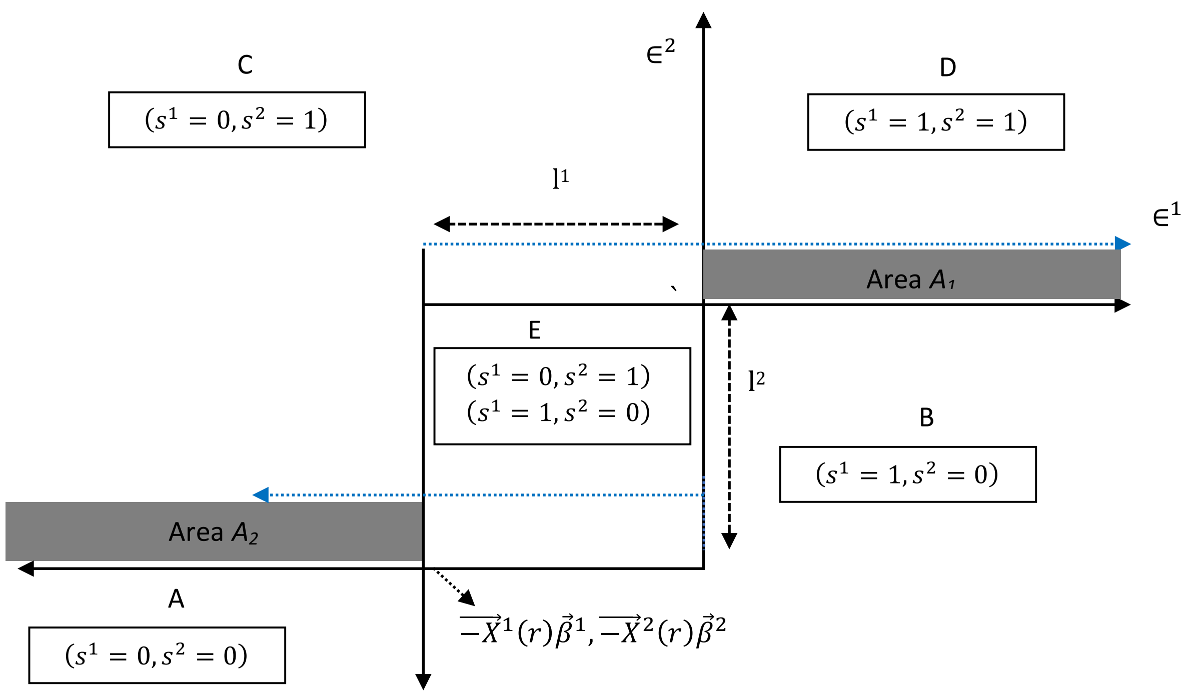


**Fig. 9 Effect on the likelihood regions of Nash-equilibria by increasing the cost of Player 2.**

through Fig. 9. This figure is constructed to analyze the outcome when the cost of Player 2 is increased. From the posterior samples we obtained that the preference parameter towards cost for Player 2 had a mean value of $-2.17$. This means Player 2 had a negative preference towards higher cost. The solid lines partitions the Two-Dimensional space

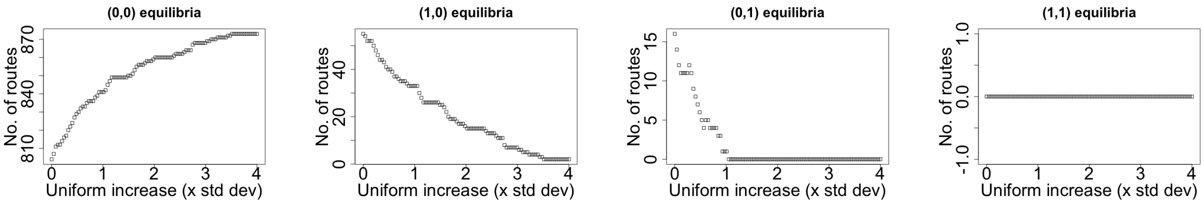


**Fig. 10 Comparing the number of routes in each Nash-equilibrium was predicted by increasing the cost of both the players.**

formed by the unknown variables $\epsilon_1$ and $\epsilon_2$ to five regions. Four of the regions have unique pure Nash equilibrium whereas the central region does not have a unique Nash equilibrium. The blue dotted line indicates the new boundaries demarcating the equilibrium regions after increasing the cost of Player 2. All the horizontal boundary lines shift upwards in the vertical direction whereas all the vertical boundary lines are unchanged. This is because the coordinates $(-\vec{X}^1\vec{\beta}^1, -\vec{X}^2\vec{\beta}^2)$ and the lengths $l^1$, $l^2$ uniquely define the regions in the Two-Dimensional space. The preference parameters for both the players ($\vec{\beta}^1$, $\vec{\beta}^2$) are unchanged as they are characteristic of the airlines. Since the cost of only Player 2 is changed, $\vec{X}^1$ also remains constant. Therefore, $\vec{X}^2$ is the only variable which defines the vertical position of the boundaries. Increasing cost shifts the boundary lines along the vertical axis ($\epsilon^2$) in the positive direction as the coefficient associated with cost for Player 2 is negative. On increasing the cost of Player 2, area corresponding to region of $(0, 0)$-Nash equilibrium increases as indicated by area $A_2$. In Fig. 8 we observe that the number of routes corresponding to $(0, 0)$-Nash equilibrium increases as expected. On the other hand, the area $A_2$ (along with some area swept by the central region) is removed from the likelihood region (region $C$ in Fig. 9) associated with equilibrium $(0, 1)$. Thus, the number of routes with $(0, 1)$ equilibrium decreases as observed from Fig. 8. Region $D$ corresponding to equilibrium $(1, 1)$ decreases in area by $A_2$ as shown in Fig. 9. $\epsilon^1$ and $\epsilon^2$ are modeled as random variables drawn from the standard normal distribution and the product of their cumulative density functions corresponding to the area of a region gives the likelihood of the equilibrium associated with that region being played. However, in the routes considered for prediction there were no routes where $(1, 1)$ equilibrium was played. The predictive model predicts this outcome with the actual values of route-attributes. Increasing cost further will further diminish the probability of $(1, 1)$ equilibrium being played closer to zero. As a result the number of $(1, 1)$ equilibrium remains at zero on increasing cost as observed in Fig. 8. The area under $(1, 0)$ equilibrium in region $B$ of Fig. 9 increases and this is seen from an increased number of routes falling in $(1, 0)$ equilibrium in Fig. 9.

Another interesting observation is that once the number routes with $(0, 1)$ equilibrium goes to zero (around 1.1 std dev in Fig. 8), the number of routes where $(1, 0)$ equilibrium is being played decreases based on the assumptions made in our model. This is because once all the routes where $(0, 1)$ equilibrium was being played converts to $(0, 0)$ or $(1, 0)$, increasing cost further does not add any additional area into $(1, 0)$ equilibrium. However, the contribution of central region decreases on increasing the cost further. The area of the central region remains the same as parameters $l^1$ and

$l^2$ are independent of the cost. However, the likelihood value associated with the same area decreases because of the nature of Gaussian density function. For example, $\phi(1.5) - \phi(1)$ is higher than $\phi(2.5) - \phi(2)$, where $\phi$ is cumulative density function of standard normal.

The interpretations through Fig. 9 were provided on increasing the cost of Player 2 only. However, if the increase was made for cost of Player 1 instead, the effects are very similar except that now the number of routes where $(1, 0)$ equilibrium is being played goes down to zero. Figure 10 shows the effect when the costs associated with both the players are increased. All equilibria except $(0, 0)$ goes down to zero eventually. Number of routes with $(0, 1)$ equilibrium drops faster compared to the routes with $(1, 0)$ equilibrium. This is because the preference parameter towards cost for Player 1 has a mean value of $-5.01$, whereas for Player 2 has a mean value of $-2.17$. Since Player 1 has a higher magnitude for the preference parameter, varying the cost has a bigger effect on the utility function of Player 1 as compared to Player 2. Therefore the number of routes by Player 1 drops more drastically towards zero as compared to Player 2.

*2. Varying market demand*

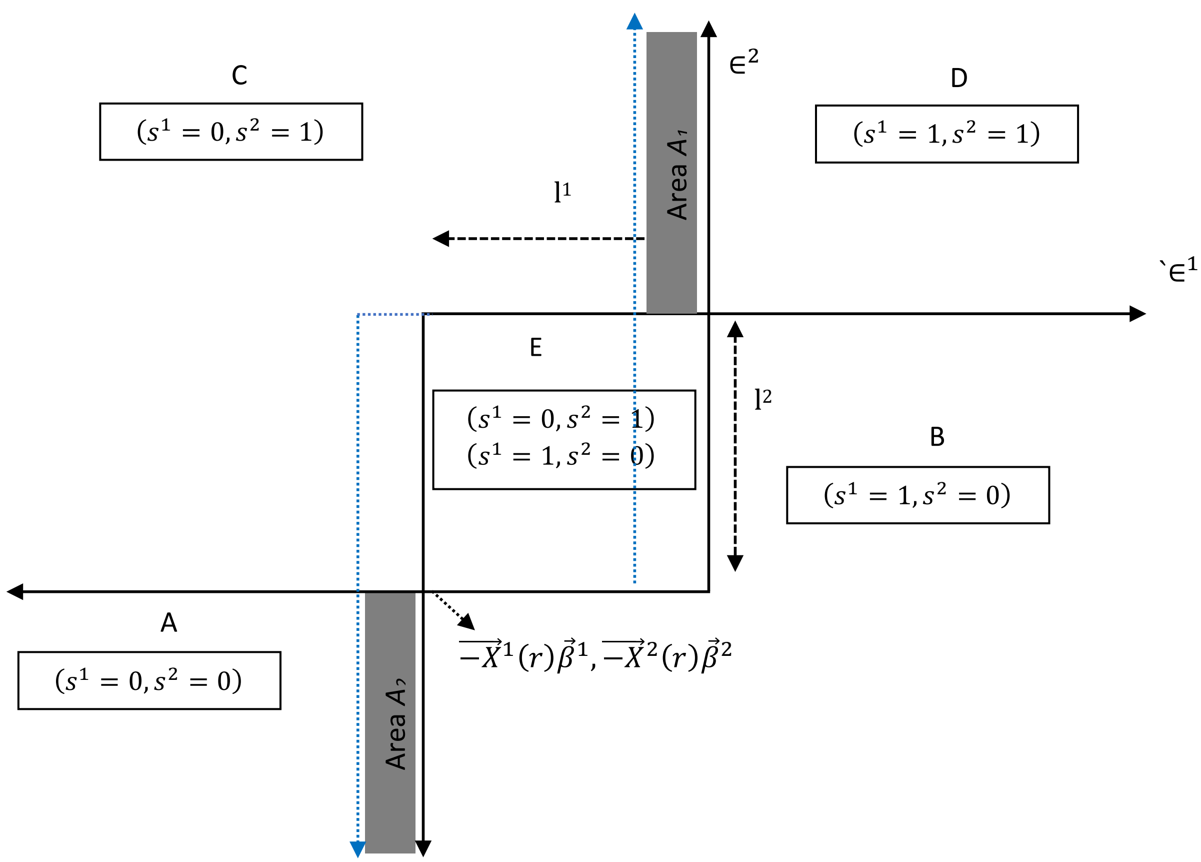


**Fig. 11 Effect on the likelihood regions of Nash-equilibria by increasing the cost of Player 2.**

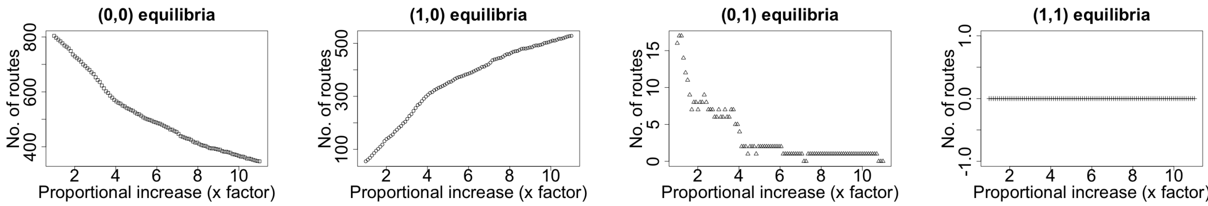


**Fig. 12 Number of routes in each Nash-equilibrium due to proportionally increasing the market demand of Player 1 (UA).**

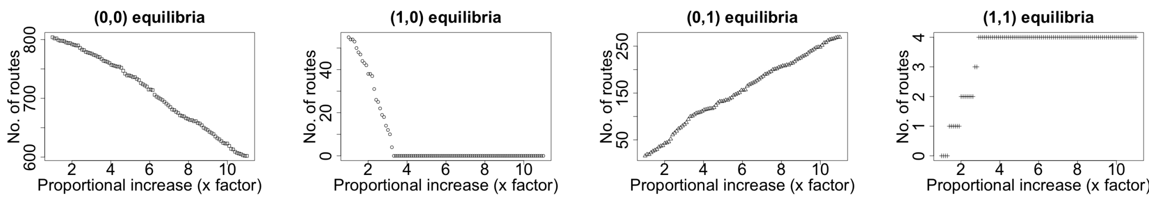


**Fig. 13 Number of routes in each Nash-equilibrium due to proportionally increasing the market demand of Player 2 (DL).**

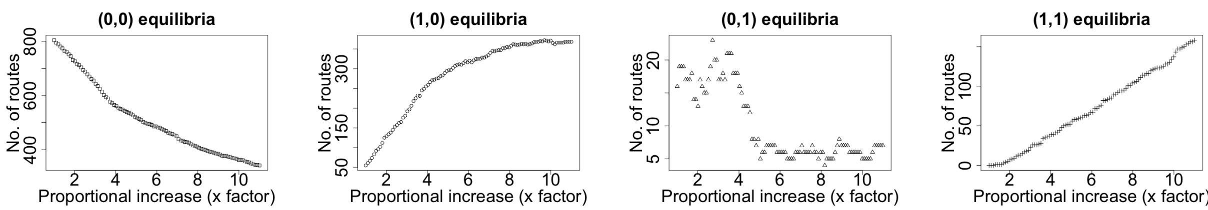


**Fig. 14 Number of routes in each Nash-equilibrium due to proportionally increasing the market demand of both the players.**

Figure 12 shows the number of routes where each of the four Nash equilibria strategies were played as a function of market demand for Player 1. The market demand for a route is increased in proportion to the original demand of that route. Figure 11 explains the effect of increasing the market demand of Player 1 on the five regions. As observed from Fig. 11, increasing market demand for Player 1 shifts all the vertical lines demarcating the boundaries of the five regions along the negative $\epsilon^1$-axis direction. The shift is in the negative direction as the preference parameters towards demand has a positive coefficient for Player 1 from the predictive distribution results. Similar to the cost, since the market demand for Player 2 is unchanged the horizontal lines are unchanged.

When market demand for Player 1 is increased, the area $A_1$ gets added to region $D$ as observed in Fig. 11. Unlike the cost where the number of routes with (1, 1) equilibrium being played remained zero, this results in a few routes with (1, 1) equilibrium. Figure 13 shows the number of routes where each of the four Nash equilibrium strategies were played as a function of market demand for Player 2. The effects are very symmetrical to the one where market demand of Player 1 was changed. Preference parameter for demand had a mean value of 3.11 for Player 1 and 0.47 for Player 2. Thus Player 1 has a relatively high preference coefficient for demand in comparison to Player 2. This result in an interesting phenomenon. Even though the number of routes where Player 1 exclusively operates or those with $(1,0)$ equilibrium decreases to zero as the market demand for Player 2 increases, Player 1 is able to compete and operate in a few routes i.e. there are a few routes with $(1,1)$ equilibrium even when market demand associated with Player 1 is increased to very high values. This was not observed when market demand for Player 1 was increased. Player 1 unanimously dominated all the routes.

Figure 13 shows the number of routes where each of the four Nash equilibria strategies were played when the market demand of both the players are increased. Unlike all the previous cases, this case results in a large number of routes with $(1,1)$ equilibrium. The number of routes with $(0,0)$ equilibrium monotonically decreases with increasing market demand. An interesting observation is that, on increasing the market demand of both players simultaneously, the number of routes with $(1,0)$ equilibrium increases whereas those with $(0,1)$ equilibrium decreases. This is because the Player 1 had a much higher preference coefficient for demand in comparison to Player 2 and therefore is able to attain higher gains from an increasing market demand in comparison to Player 2. This results in a large number of routes where only Player 1 operates.

For a policy maker, modeling the decisions of airlines on network structure will allow them to form policies that encourage desirable features of the ATN. From policy designers' standpoint, the only variable that can be directly influenced is the cost. Non-stop distance is a geographic factor, which is fixed for every route. Demand is governed more by market factors and is difficult to be directly influenced. Through indirect approaches such as installing a cheap and fast alternate mode of transportation between the origin and destination can be utilized, demand still is much harder to directly influence. Cost is relatively easier to directly influence by levying taxes, imposing penalties for entering and operating service on congested routes, and airport regulation (such as single-till or dual-till [31]). For example, a policy maker may be interested in encouraging an increase in flights to smaller airports. In this case, the policy maker can setup economic incentives to drive operations towards the smaller airports. Our proposed model can help analyze the effects of such policy initiatives on airline decision-making and allow for comparison of different policies on the air transportation system.

The evolutionary model developed here can be used to study how the network evolution can be influenced by influencing the decision parameters. We conducted simulation experiments by varying the cost and demand values of each route. The number of routes has direct implications on network properties such as robustness, resilience, and connectivity.

## IX. Conclusion

We modeled competition between airlines at the route level as discrete games of perfect information where the parameters of the model are estimated using an approach based on airport presence. The airport presence based approach is an improvement over existing approaches based on Gibbs sampling as it removes problems such as over-fitting. This is reflected from the route prediction accuracy of the model.

The proposed approach that considers competition improved the prediction accuracy over the discrete choice based approach that ignores competition, on routes where either a new route was added or the existing route was deleted. However, the overall accuracy was higher in discrete choice based approach as compare to the proposed approach. This is because the discrete choise based approach considers existing route structure to make predictions for the next period, but the proposed approach does not make use of existing route structure. The proposed approach predicts considering only the route-attributes. A discrete choice based approach achieves better overall accuracy at the cost of predicting changes on routes where changes did not actually occur. The goal of the proposed approach is to capture the decision making behavior of the airlines. The approach aims to provide guidelines to policymakers on the effect of their policies on route addition and route deletion. From this perspective, the model performs better than a discrete choice based approach as it predicts the routes where changes did actually occur with a higher accuracy than a discrete choice based approach.

For two major airlines, cost was the most important route parameter while making routing decisions. From the preference estimates it was observed that increasing the cost decreased the probability of a route getting added, whereas increasing the market demand increased the probability of the route getting added. These effects were further validated using forward simulations. Forward simulations provided insights about the Nash equilibrium strategies on varying the route characteristics.

A limitation of the approach is that it does not take into account the effect of the decision made on a route in the current period on the decision made on other routes. This modeling decision was based on the fact that addition or deletion of routes happens only in less than 10% of the routes from one period to the next. This is evident from the accuracy of prediction of the model. Although the overall accuracy of the discrete choice model is higher, the discrete games-based approach is more accurate in predicting the airline decision routes where an addition or deletion was actually made. The latter accuracy is more important for the objective of the model which is to guide policymakers to understanding how airlines make route selection decisions.

## Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation through NSF CMMISYS grant 1360361 and Prof. Ilias Bilionis and Apoorv Maheshwari at Purdue University.

## References

[1] Belobaba, P., *The airline planning process*, Wiley West Sussex, UK, 2009. doi:10.1002/9780470744734.

[2] Kushige, T., "A solution of most profitable aircraft routing," *AGIFORS Symposium Proceedings*, Vol. 3, 1963, p. 1963.

[3] De Lamotte, H., and Mathaisel, D., "Experience with MPSX-MIP for Olympic Airways fleet assignment problems," *MIT Flight Transportation Laboratory Memorandum*, 1983.

[4] Kotegawa, T., "Analyzing the evolutionary mechanisms of the air transportation system-of-systems using Network Theory and machine learning algorithms," Ph.D. thesis, Purdue University, 2012.

[5] Evans, A., and Schäfer, A. W., "Simulating airline operational responses to airport capacity constraints," *Transport Policy*, Vol. 34, 2014, pp. 5–13. doi:10.1016/j.tranpol.2014.02.013.

[6] Sha, Z., Moolchandani, K., Panchal, J. H., and DeLaurentis, D. A., "Modeling Airlines' Decisions on City-Pair Route Selection Using Discrete Choice Models," *Journal of Air Transportation*, Vol. 24, No. 3, 2016, pp. 63–73. doi:10.2514/1.d0015.

[7] Doyme, K., Dray, L., O'Sullivan, A., and Schäfer, A., "Simulating Airline Behavior: Application for the Australian Domestic Market," *Transportation Research Record*, 2019, p. 0361198119826533. doi:10.1177/0361198119826533.

[8] Bresnahan, T. F., and Reiss, P. C., "Empirical models of discrete games," *Journal of Econometrics*, Vol. 48, No. 1-2, 1991, pp. 57–81. doi:10.1016/0304-4076(91)90032-9.

[9] Ciliberto, F., and Tamer, E., "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, Vol. 77, No. 6, 2009, pp. 1791–1828. doi:10.3982/ecta5368.

[10] Chib, S., and Greenberg, E., "Understanding the metropolis-hastings algorithm," *The american statistician*, Vol. 49, No. 4, 1995, pp. 327–335. doi:10.1080/00031305.1995.10476177.

[11] Berry, S. T., "Airport presence as product differentiation," *The American Economic Review*, Vol. 80, No. 2, 1990, pp. 394–399.

[12] Jaillet, P., Song, G., and Yu, G., "Airline network design and hub location problems," *Location science*, Vol. 4, No. 3, 1996, pp. 195–212. doi:10.1016/S0966-8349(96)00016-2.

[13] Balakrishnan, A., Chien, T. W., and Wong, R. T., "Selecting aircraft routes for long-haul operations: a formulation and solution method," *Transportation Research Part B: Methodological*, Vol. 24, No. 1, 1990, pp. 57–72. doi:10.1016/0191-2615(90)90032-T.

[14] Lohatepanont, M., and Barnhart, C., "Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment," *Transportation Science*, Vol. 38, No. 1, 2004, pp. 19–32. doi:10.1287/trsc.1030.0026.

[15] Dantzig, G. B., "Linear Programming and Extensions. Princeton landmarks in mathematics and physics," *Princeton University Press*, 1963.

[16] Song, K., Lewe, J.-H., and Mavris, D., "A Multi-Tier Evolution Model of Air Transportation Networks," *14th AIAA Aviation Technology, Integration, and Operations Conference*, 2014, p. 3267. doi:10.2514/6.2014-3267.

[17] Boguslaski, C., Ito, H., and Lee, D., "Entry patterns in the southwest airlines route system," *Review of Industrial Organization*, Vol. 25, No. 3, 2004, pp. 317–350. doi:10.1007/s11151-004-1970-5.

[18] Sha, Z., Moolchandani, K. A., Maheshwari, A., Thekinen, J., Panchal, J., and DeLaurentis, D. A., "Modeling Airline Decisions on Route Planning Using Discrete Choice Models," *15th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2015. doi:10.2514/6.2015-2438.

[19] Moolchandani, K., Sha, Z., Maheshwari, A., Thekinen, J., Davendralingam, N., Panchal, J., and DeLaurentis, D. A., "Hierarchical Decision-Modeling Framework for Air Transportation System," *16th AIAA Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics, 2016. doi:10.2514/6.2016-3154.

[20] Berry, S. T., "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, 1992, pp. 889–917.

[21] Sugawara, S., and Omori, Y., "Duopoly in the Japanese airline market: bayesian estimation for the entry game," *The Japanese Economic Review*, Vol. 63, No. 3, 2012, pp. 310–332. doi:10.1111/j.1468-5876.2011.00545.x.

[22] "Air Carriers: T-100 Domestic Market (U.S. Carriers)," `http://www.transtats.bts.gov/Fields.asp?Table_ID=258`, ???? Accessed: 2017-10-11.

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

[24] Levine, M. E., "Airline competition in deregulated markets: theory, firm strategy, and public policy," *Yale J. on Reg.*, Vol. 4, 1986, p. 393.

[25] "Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports)," `https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/`, ???? Accessed: 2017-10-11.

[26] Train, K. E., *Discrete choice methods with simulation*, Cambridge university press, 2009. doi:10.1017/CBO9780511805271.

[27] Ben-Akiva, M. E., and Lerman, S. R., *Discrete choice analysis: theory and application to travel demand*, Vol. 9, MIT press, 1985.

[28] "Air Carrier Financial: Schedule P-5.2," `http://www.transtats.bts.gov/Fields.asp?Table_ID=297`, ???? Accessed: 2017-10-11.

[29] Bureau, U. C., "Statistical abstract of the United States: 2012," *Commerce USDo, editor*, 2012, p. 111.

[30] Hanley, J. A., and McNeil, B. J., "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology*, Vol. 143, No. 1, 1982, pp. 29–36. doi:10.1148/radiology.143.1.7063747.

[31] Czerny, A. I., "Price-cap Regulation of Airports: Single-till Versus Dual-till," *Journal of Regulatory Economics*, Vol. 30, No. 1, 2006, pp. 85–97. doi:10.1007/s11149-006-0010-9.