
AIRLINE CHOICE FOR DOMESTIC FLIGHTS IN SÃO PAULO METROPOLITAN AREA: AN APPLICATION OF THE CONDITIONAL LOGIT MODEL

Marcelo Baena Moreno
Energycarc Industrial Limitada
São Paulo, Brazil

ABSTRACT

Using the conditional (multinomial) LOGIT model, this paper addresses airline choice in the São Paulo Metropolitan Area. There are two airports in this region, where two, three or even four airlines compete for passengers flying to an array of domestic destinations. The airline choice is believed to be a result of the tradeoff passengers face among flight cost, flight frequency and airline performance. It was found that the lowest fare better explains airline choice than the highest fare, whereas direct flight frequencies give better explanation to airline choice than indirect (connections and stops) and total (direct plus indirect) ones. Out of 15 variables tested, the lowest fare was the variable that best explained airline choice. However, its signal was counterintuitive (positive) possibly because the cheapest airline was offering few flights, so passengers overwhelmingly failed to choose the cheapest airline. The model specification most adjusted to the data considered the lowest fare, direct flight frequency in the travel day and period (morning or afternoon peak) and airline age. Passengers departing from São Paulo-Guarulhos International Airport (GRU) airport make their airline choice in terms of cost whereas those from São Paulo-Congonhas Airport (CGH) airport do not. Finally, senior passengers place more importance on airline age than junior passengers.

Marcelo Baena Moreno has a B.Sc. degree in Civil Engineering from the University of Campinas, Brazil, a M.Sc. degree in Air Transport and Airports from the Instituto Tecnológico de Aeronáutica, Brazil, and is currently pursuing a Ph.D. degree in Air Transport at Cranfield University, United Kingdom. His research interests focus on airport and airline choice in competitive markets. In addition, he has been working for several years as a General Manager for Energycarc Industrial Limitada, an importing/exporting/distributing company in Brazil.

INTRODUCTION

Despite incidents related to air transport security such as the terrorist attacks on September 11, 2001, and seasonable widespread infectious diseases such as SARS in China and elsewhere, air transport demand has a long-term rising trend as a result of world population increase and industry development.

As a result of deregulation in many countries worldwide, airlines are now facing a different competitive condition, with new airline entrants; some of them employing a low-cost/low-fare strategy that has changed air travel from expensive and elitist to more affordable and for a wider population.

In this scenario, airlines are searching eagerly to enhance their market share not only in the routes they already operate but also in new potential markets. Therefore the way passengers choose airlines for their desired flights constitutes both crucial information for the airlines' strategic plans and a relevant subject of research for transportation engineers.

To describe the airline choice process, the majority of researchers have used at least a variable accounting for the cost of the flight, generally the fare, and another variable accounting for the flight frequency. This paper goes further by using a variable of airline performance.

This paper aims at determining from a set of candidates the variables that have the best explanatory power on airline choice made by the passengers whose travel starts in the São Paulo Metropolitan Area. This region is well served by two airports: São Paulo Guarulhos International Airport (GRU) and São Paulo-Congonhas Airport (CGH), which are outstanding countrywide in terms of embarked and disembarked passengers.

This paper extends the research on airline choice by presenting and discussing results achieved in the analysis of airline choice in other regions of the world and bringing about results for the São Paulo Metropolitan Area.

BACKGROUND

Prousaloglou and Koppelman (1995) analyzed airline choice made by passengers originating in Dallas and Chicago, in the US. Multinomial LOGIT was used as a modeling tool, and the passenger market was segmented according to travel frequency, travel purpose and experience with traveling to different destinations. It was found that the probability of choosing a carrier increases with an increase in the level of service (share of the carrier flights in the origin-destination city pair), the square of this variable has a negative signal, the effect of origin market presence is positive but unexpectedly not significant, frequent flyer program (FFP) membership and most active membership are positive and highly significant, the carrier's attractiveness and its market share are positively affected by FFP

membership, the most active membership reflects incremental effect of participation in a FFP, therefore those who actively participate are more likely to choose the carrier of this program than those who just participate, showing the loyalty-inducing effect of FFP membership. Finally, relevant scenarios were built. The carrier choice probability increases from 50% to 72% for travelers who become members of that carrier FFP and to 92% for frequent travelers who actively participate in that carrier's FFP.

Yai, Takada, and Okamoto (1997) examined the travel characteristics of international passengers traveling to the Asian region as well as their choice of air carrier for international flights at Tokyo New International Airport (Narita) in Japan, using ordered LOGIT model. Residents in Central America traveling on sightseeing were the ones who visited more countries in the current trip to Asia, whereas residents in East Asia were those who visited fewer countries. Japan, Singapore and Hong Kong are considerably used as transit ports for passengers traveling elsewhere within Asia. Moreover, would a travel pattern be defined, it would involve visiting countries located near each other. Regarding parameter estimations for passengers preferring economy class, for the US, Canada and Europe passenger signals were intuitive for fare (negative), travel time (negative), frequency (positive) and airline nationality (positive). For passengers from Korea, China and Southeast Asia, airline nationality was negative. For travelers from Southeast Asia time was positive. US passengers demonstrated the highest willingness to pay for national carrier, whereas Canadian passengers placed the highest value on travel time and finally Southeast Asian passengers placed the highest importance on flight frequency.

Pels, Nijkamp, and Ritveld (1998) studied the conjoint choice of airport and airline in the San Francisco Bay Area, using a nested LOGIT model, building two situations of sequential choice: (a) first airport choice and then airline choice; and (b) first airline choice, followed by airport choice. There was not an expressive difference in the estimations of the utility function between business and non-business passengers. These parameters seemed to vary more across time than across market segments. Anyway, they concluded that the estimated parameters were rather robust. Moreover, they concluded that airport choice happens first, and then airline choice, with them not being simultaneous choices.

Using an aggregate-level Markovian type model, Suzuki (2000) proposed a method to model the relationship between on-time performance and market share in the airline industry. The model incorporates the idea that passengers who experienced flight delays are more likely to switch airlines in their subsequent flights than those that did not face delays. A delay was considered if it surpassed 15 minutes from schedule time. The paper concludes that on-time performance affects a carrier's market share primarily

through the passengers' experience not through the advertisement of performance.

Mason (2001) analyzed business travel decision making within the UK, interviewing both individual travelers and their corporate travel managers. Eighty percent (80%) of the companies used only one travel agent. Eighty-five percent (85%) of the travelers and their travel managers used phone calls as a booking channel. The traveler selects his or her own flights 52% of the time, the traveler's secretary selects it 25% of the time and the travel department makes the selection 23% of the time. While travel managers think corporate travel policies (CTP) make travel easier, travelers are less convinced about this. Unlike travelers, travel managers think that CTP reduce traveler uncertainty. Travel managers disagree that CTP put a constraint on travel planning, while travelers were neutral. Both travelers and travel managers agree that CTP reduce travel choice, furthermore travel managers agree more strongly than travelers that CTP save the company money. Forty-seven percent (47%) of the travel managers see consolidated spending with one alliance as beneficial. In addition, 37% of them see that alliance development has a neutral effect on the company travel expenditure. Sixty-five (65%) of travel managers have a positive attitude towards low cost airlines, whereas only 32% of travelers do. Price is clearly seen by both travelers and travel managers as the main advantage of low cost airlines. Finally, 70% of the travel managers and travelers believe that video conference technology and the Internet did not have a substantial effect on the number of trips taken.

Hensher (2001) contributed to the literature of discrete choice models by considering structures for the specification of unobserved effects in the utility function. Using data from the non-business market for the Sydney-Canberra corridor served by car, the conclusion was that past research has under valued travel time savings.

Suzuki and Walter (2001) presented a framework that investigates how frequent flyer miles can be used in the most effective way to reduce air travel costs by companies that are considering the use or are already using mileage redemption strategies. Among three candidate methods, the conclusion is that the mileage optimization method is the best one, followed respectively by the lowest fare redemption method and the lowest fare method.

Armstrong, Garrido, and Ortúzar (2001) studied the choice of urban trips in Chile. Although it does not analyze airline choice, this paper contributes to the literature of discrete choice models in the sense that it focus on the subjective value of time (SVT), which is the marginal rate of substitution between travel time and cost. Since the SVT point estimate follows an unknown distribution a priori, this paper proposes two forms for building confidence intervals for a certain probability level: the t-test and the LR-test, constructing Multinomial Logit, Hierarchical Logit and Box-Cox

Logit. The conclusions are that the interval's mid-point is greater than the SVT point estimate, and smaller confidence intervals should be derived from more significant parameters. Both the t-test and the LR-test provided an easy and practical way to obtain good confidence intervals for the SVT. Finally, as trip maker income increases, the SVT point estimates also grow, but the variation of the intervals' mid-point is much more drastic and the range of values increases considerably.

Using ANOVA, Yoo and Lee (2002) studied airline choice for international flights made by Korean air passengers departing from Incheon International Airport, which is an off-shore airport that serves Korea's capital, Seoul. The most important airline service attributes were, respectively, air fare, convenience of flight schedule, on time performance, and seat availability. People who have less than a college education placed higher importance on in-flight service. Travelers in their thirties or forties, and individual enterprisers placed higher importance on air fare. Passengers in their twenties and fifties, professionals and individual enterprisers, and less educated people placed higher importance on tour information and extra service from airlines. Travelers with higher income, professionals, passengers with less than a college degree, and those participating in group tours and people traveling more than 11 times a year placed higher importance on reputation and image of airlines. Passengers with middle income and office workers placed higher importance on safety. Females and older travelers placed higher importance on recommendations and experiences. When travelers and relatives paid for the ticket, they placed more importance on safety. Finally, business travelers and those visiting friends and relatives placed more importance on scheduling and on-time performance.

Turner (2003) analyzed the profile and airline choice of passengers departing from London Gatwick Airport in England to Amsterdam Schiphol Airport in the Netherlands. Passengers of two airlines were surveyed, EasyJet (EZ, a no-frills carrier) and British Airways (BA, a network carrier). EZ flyers fly mostly on leisure, are younger, come from a diversity of occupations, do not participate so much in FFP and are less frequent flyers, whereas BA passengers fly mostly on business, are older, are businessmen, participate in FFP and are extremely frequent flyers. Ninety-seven (97%) of EZ passengers rated price as important, 75% indicated flight timings and 33% said frequency. Eighty-five (85%) of BA passengers rated flight timings as important, 33% did not know how much the ticket cost, 26% rated FFP points as important, ahead of reliability/punctuality (25%) and frequency (17%). Regarding airline choice, 47% of EZ passengers considered another carrier for the trip, while 44% of BA passengers did. The trip purpose influenced the access mode: business travelers accessed the airport by taxi whereas leisure passengers accessed by bus/coach or train.

Finally, some EZ passengers rated 30 pounds as more than expected for the ticket price while others rated 60 pounds as a lot less than expected.

Lijesen (2006) conducted a stated preference survey with Dutch respondents, who were exposed to 16 choice problems each. These choices mimic a trip from Amsterdam to New York. Estimating a mixed logit model, it was found that westbound long-haul leisure passengers in general prefer flights with afternoon arrivals and that the majority of these travelers prefer arriving before their desired arrival time than arriving after their desired arrival time, implying that flights should not be spaced equally over time, but be biased towards arriving earlier.

METHODOLOGY

The LOGIT model has been the most widely used approach to cope with multiple-choice situations in transportation engineering, especially in the majority of the papers analyzed in the previous section. To build the LOGIT model, some considerations related to the passengers' choice process are imperative.

Each passenger presents a consistent structure of preferences, based on the utility each alternative choice can provide, in a way that the passenger chooses the option (airline) whose utility is the maximum among the available choices. This choice behavior can be expressed mathematically by the following equation:

$$U_{in} \geq U_{jn} \text{ for all } j, 1 \leq j \leq z \quad (1)$$

Where: U_{in} is the utility that passenger n obtains by choosing airline i , U_{jn} is the utility that passenger n obtains by choosing airline j , z is the number of airlines (alternatives) available for choice.

Since the perception of the attributes that each alternative offers may vary widely from passenger to passenger, and even the characteristics usually measured being constant for two different passengers, the utility of each alternative airport is not regarded from the same standpoint, therefore it is wise to include a random element to the travel choice, that is added to the deterministic one, forming the theoretical basis for the stochastic choice. The stochastic formulation of the utility function is expressed as:

$$U_{in} = V_{in} + \varepsilon_{in} \text{ for all } i, 1 \leq i \leq z \quad (2)$$

Where: U_{in} is the utility that passenger n obtains by choosing airline i , V_{in} is the deterministic part of the utility function for alternative i chosen by passenger n , ε_{in} is the random part of the utility function for alternative i

chosen by passenger n , z is the amount of choices considered available for passenger n .

The LOGIT model assumes that the random components of the utility function are independent and identically distributed with a Gumbel function (double exponential) as Kanafani (1983) explains. The probability function that denotes the choice of an alternative made by one passenger is given by:

$$p_{in} = \frac{e^{V_{in}}}{\sum_{j=1}^j e^{V_{jn}}} \quad (3)$$

Where: p_{in} is the probability of passenger n choosing alternative i (each alternative is an airline in this paper), among the j alternatives (airlines); V_{in} is the deterministic part of the utility function of alternative (airline) i .

V_{in} enhances parameters α_k whose estimation has been accomplished using NLOGIT 3.0 (Econometric Software, Inc., 2002). Multinomial LOGIT models are classified as follows: (a) models whose variable values are input the same across all alternatives for the same observation (passenger), as they are individual characteristics; and (b) models whose variable values are attributes of the alternatives (perceived by passengers), and variable values that remain constant across alternatives (for the same passenger) are also allowed.

The latter is the model that this paper employs, also known as the conditional LOGIT model, which estimates variable parameters using the Maximum Likelihood Method. For the iterations, the Newton Method was used since it produced quick convergence for most calibrated models. As a measure of goodness-of-fit, the average probability of a correct prediction was generated.

Sampling

Although the São Paulo Metropolitan Area groups several towns, seven of them (São Paulo, Guarulhos, Santo Andre, São Bernardo do Campo, São Caetano do Sul, Diadema and Osasco) have been chosen to represent the trip origins in this region, because of two reasons: (a) they represent 79% of the electric power consumption in the region; and (b) the data was primarily collected for an airport choice experiment, and in that case the calculation of the access time from the other towns was not likely to lead to sound values. For airport choice analysis we used Moreno and Muller (2003, 2004).

The analyzed airports were São Paulo-Congonhas Airport (CGH), located in São Paulo) and São Paulo-Guarulhos International Airport (GRU),

located in Guarulhos, a city neighboring São Paulo. The criteria for destination selection were that: (a) there must have been departures to these destinations from both airports; (b) flights supplied by plural airlines; and (c) observed traffic must have surpassed 100,000 passengers. This number, as Windle and Dresner (1995) explain, prevents small sample bias that is usually associated with less popular destinations.

The first and third requisites were evaluated through the last statistical report of the Department of Civil Aviation available to the date of data collection, the report of the year 2000. If this survey had been designed only for airline choice analysis, the first requisite could have been waived. The second requisite was evaluated through airlines' websites.

Therefore 21 airports (corresponding to 19 cities, since there are 2 multiple-airport destinations) were studied in this paper, as follows: (a) BPS (Porto Seguro); (b) BSB (Brasilia); (c) CGR (Campo Grande); (d) CNF (Belo Horizonte); (e) CWB (Curitiba); (f) FLN (Florianopolis); (g) FOR (Fortaleza); (h) GIG (Rio de Janeiro); (i) GYN (Goiania); (j) IGU (Foz do Iguacu); (k) JOI (Joinville); (l) LDB (Londrina); (m) NVT (Navegantes); (o) PLU (Belo Horizonte); (p) POA (Porto Alegre); (q) RAO (Ribeirao Preto); (r) REC (Recife); (s) SDU (Rio de Janeiro); (t) SSA (Salvador); (u) UDI (Uberlandia); and (v) VIX (Vitoria).

The number of competing airlines ranged from 2 to 4 depending on the destination, herewith denoted by Airline 1, Airline 2, Airline 3 and Airline 4.

The passenger profile was obtained by revealed preference (RP) survey carried out at the departing lounges of GRU and CGH during the weekdays of two consecutive weeks [February 18 to March 1, 2002, during the peak hours of access to airports, i.e., from 7:00 a.m. to 10:00 a.m. (morning peak) and from 5:00 p.m. to 8:00 p.m. (afternoon peak)].

Since these data were collected primarily for an airport choice experiment, these periods were chosen because the average vehicle speeds in São Paulo have been measured during these peak periods by CET, a traffic engineering company, enabling the calculation of access time to the airports.

Aiming at a maximization of the explanatory power of the collected data and a minimization of time and cost of data collection, compilation and analysis, 1,923 passengers were interviewed: 897 at GRU and 1,026 at CGH. This amount of observed data has been considered satisfactory taking into account Koppelman and Chu (1985) who calculated the amount of observations required for relatively simple disaggregate choice models. However, some observations have been excluded for the airline choice analysis. The passengers from a fifth airline were removed from this analysis because the ticket fare could not be obtained for the period in question. This did not pose a problem because this airline had few flights and covered few destinations. The passengers whose declared airline operated only part of the itinerary to the chosen destination (e.g., one leg), but not it completely, were

removed because the flight frequency could not be input. Therefore the final number of observations for the airline choice experiment was slightly less than that for the airport choice experiment, 1,900 passengers.

The literature tends to classify the passenger market in a way that enables inferences on the airline choice made by homogeneous passenger segments. Table 1 presents the results of the interviews according to market segmentation criteria.

Table 1. Results of the interview with passengers

Sample segmentation criteria	Number of passengers	Passenger market segments
Period of departure	844	Morning peak
	1056	Afternoon peak
Airport of departure	879	GRU
	1021	CGH
Travel purpose	1387	Business
	513	Non-business
Place of residence	926	Residents
	974	Visitors
Passenger age	963	Junior (up to 36 years old)
	937	Senior (over 36 years old)
Household monthly Income	357	Lower income (up to R\$ 3k)
	1150	Middle income (between R\$ 3k and R\$ 10k)
	393	Higher income (over R\$ 10k)
Flying frequency (departures from CGH and GRU in the previous year)	385	Occasional flyers (up to 1 flight)
	502	Fairly frequent flyers (between 2 and 6 flights)
	1013	Flyers extremely frequent (over 6 flights)
Access mode	100	Car ride-and-park (paid parking)
	716	Car ride-and-kiss
	864	Taxi
	123	Bus
Flight duration	793	Short-haul flights (up to 1 hour)
	1107	Long-haul flights (over 1 hour)
Air carrier	817	Airline 1
	615	Airline 2
	445	Airline 3
	23	Airline 4
Proximity to airports (1)	392	Closer to GRU
	1436	Closer to CGH
Proximity to airports (2)	954	Extremely closer to one airport
	388	Rather closer to one airport
	178	Fairly equidistant to both airports
Popularity of the destination	331	Flying to the most popular destination
	1569	Flying to other destinations
Loyalty to airports	435	Loyal to CGH
	1020	Disloyal to both airports
	247	Non-experienced with airports
	198	Loyal to GRU

Note. GRU = São Paulo-Guarulhos International Airport; CGH - São Paul-Congonhas Airport

Variable selection

Three types of variables were chosen to be tested: (a) those associated with flight cost; (b) those related to flight frequency; and (c) those associated with airline performance.

Using the conditional LOGIT model, the utility function of an alternative was designed as the summation of the effects of the variables pre-multiplied by a parameter whose estimation is one of this paper's goals. The model built was abstract, for example, the coefficients of the variables were the same for all alternative airlines.

According to the flight destination, each passenger n , $1 \leq n \leq 1900$, has been represented by two, three or even four generic decision functions. AIRLINE was a variable denoting the airline in question, ranging from 1 to 4. COUNTER was a variable denoting the amount of airlines available for choice for each passenger, ranging from 2 to 4. Listed below are the utilities for a passenger who could choose among the four airlines:

$$V_{A1n} = \alpha_1 COST_{A1n} + \alpha_2 FREQUENCY_{A1n} + \alpha_3 PERFORMANCE_{A1n} \quad (4)$$

$$V_{A2n} = \alpha_1 COST_{A2n} + \alpha_2 FREQUENCY_{A2n} + \alpha_3 PERFORMANCE_{A2n} \quad (5)$$

$$V_{A3n} = \alpha_1 COST_{A3n} + \alpha_2 FREQUENCY_{A3n} + \alpha_3 PERFORMANCE_{A3n} \quad (6)$$

$$V_{A4n} = \alpha_1 COST_{A4n} + \alpha_2 FREQUENCY_{A4n} + \alpha_3 PERFORMANCE_{A4n} \quad (7)$$

Listed below is the choice probability of airline 1 (A1) for this passenger:

$$p_{A1n} = \frac{e^{V_{A1n}}}{\sum_j e^{V_{jn}}} \quad (8)$$

Where: COST is a variable associated with the flight cost; FREQUENCY is a variable associated with the flight frequency; PERFORMANCE is a variable related to the airline performance in the market; p_{A1n} is the probability that passenger n chooses airline 1 (A1); α_k is the parameter (coefficient) related to the variable k , being $k = 1$ for COST, $k = 2$ for FREQUENCY and $k = 3$ for PERFORMANCE.

Variables associated with cost

It is unconceivable to model airline choice without a variable associated with flight cost, since the fare is part and parcel of an airline's marketing strategy. In this paper the lowest fare (LFARE) and the highest one (HFARE) were tested, their values expressed in Brazilian currency [Real (R\$)] and obtained from Panrotas (2002a, 2002b).¹

Variables associated with flight frequency

To portray the airlines' level of service, twelve variables of flight frequency have been tested. These variables were built in terms of the following criteria: (a) the existence of connections or stops (direct flights, indirect flights and the sum of the two); (b) the travel period (morning peak or afternoon peak); and (c) the day of the week. In terms of the second criterion, the passengers were interviewed at the moments prior to their departure, either at the check-in lounge or at the waiting lounge. The morning peak was considered from 7:00 a.m. to 10:00 a.m. and the afternoon peak from 5:00 p.m. to 8:00 p.m. The flight frequencies across periods of the day and across days of the two weeks when the interviews took place were determined through the websites of the airlines that offer regular flights and operate at the analyzed airports. Only flights available at the chosen airport were considered for each passenger. Although the interviews had taken place during the weekdays, weekend flight frequency was also accounted for since it increases the utility associated with the alternative airline.

For each of the built variables of frequency, its value was collected for the chosen airline and the airlines not chosen, using the following variables:

1. DDPF: Direct flight frequency in the travel day and period;
2. DDF: Direct flight frequency in the travel day;
3. DPF: Direct flight frequency in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled;
4. DWF: Direct flight weekly frequency irrespective of day and period;
5. IDPF: Indirect flight (with connections or stops) frequency in the travel day and period;
6. IDF: Indirect flight frequency in the travel day;
7. IPF: Indirect flight frequency in the travel period in all days of the week when the passenger traveled;
8. IWF: Indirect flight weekly frequency irrespective of day and period;
9. TDPF: Total flight (direct plus indirect) frequency in the travel day and period;

¹ The exchange rate to US dollars at the time of the survey was US\$ 1.00 = R\$ 2.50.

10. TDF: Total flight frequency in the travel day;
11. TPF: Total flight frequency in the travel period in all days of the week when the passenger traveled; and
12. TWF: Total weekly flight frequency irrespective of day and period.

Variables related to airline performance

It is recognized that there is a myriad of variables that serve as candidates to represent airline performance. However, some of them require a specific question in the RP survey, such as rating airline overall image according to each passenger. Since the sample of this paper was primarily collected for an airport choice experiment, variables requiring a question could not be tested. To portray the airline performance, the airline age (AGE) was tested. This variable is easy to get even after the passenger survey took place, and it represents the number of years the airline has been in the market. This variable is able to portray recognition, image or simply market habit.

Considerations for the models

The value of the variables was input directly in the decision function, without any mathematical modification, enabling the immediate analysis of the tradeoffs between the variables pertaining to the same model (what happened in the models with 3 variables). To begin with, 39 models were calibrated. Variables belonging to the same category did not take part of the same model.

Supposing a model considering three variables, Table 2 presents the values of these variables, which a fictitious passenger may have faced. It is also supposed that he or she flew Airline 1 from CGH to PLU, departing in the afternoon peak of February 25, 2002.

Table 2. Variables of lowest fare, direct flight frequency in the travel day and period and airline age for a fictitious passenger

CHOICE	AIRLINE	COUNTER	LFARE	DDPF	AGE
1	1	4	372	3	75
0	2	4	480	3	41
0	3	4	337	1	69
0	4	4	184	1	1

RESULTS**Models containing one explanatory variable**

Fifteen models belonging to this category were built, using one by one the 15 variables selected in the previous section of this paper. The comparison among these models brings out the variable with the best explanatory power on airline choice in the São Paulo Metropolitan Area. Table 3 presents the calibration results of these models.

The signals for the coefficients were positive as expected in the case of FREQUENCY and PERFORMANCE. Indeed a higher supply of flights and a longer airline are desired and their increase increases airline choice. However, in the case of COST the signals were positive while negative was expected. A possible explanation for this outcome is that the cheapest airline was offering few flights, so passengers overwhelmingly failed to choose the cheapest airline.

Table 3. Models with one explanatory variable

Model	Variable	Coefficient	t-statistic	Average probability of correct prediction	Average probability of the alternative not chosen
1	LFARE	0.0059	16.531	0.3609	0.2791
2	HFARE	0.0020	10.083	0.3307	0.2923
3	DDPF	0.2547	11.620	0.3364	0.2898
4	DDF	0.0530	10.821	0.3341	0.2908
5	DPF	0.0395	11.445	0.3358	0.2900
6	DWF	0.0090	10.971	0.3349	0.2904
7	IDPF	0.0808	7.850	0.3240	0.2952
8	IDF	0.0164	7.318	0.3224	0.2959
9	IPF	0.0135	7.904	0.3241	0.2951
10	IWF	0.0030	7.567	0.3229	0.2957
11	TDPF	0.0841	10.814	0.3320	0.2917
12	TDF	0.0174	10.303	0.3297	0.2927
13	TPF	0.0138	10.817	0.3321	0.2917
14	TWF	0.0031	10.607	0.3307	0.2922
15	AGE	0.0183	15.200	0.3497	0.2840

The t-Student statistics were satisfactory, presenting a modulus higher than 2, whereas the null p-value in all the cases also indicated satisfactory participation of the variables in the models. The calibration of the models with one variable revealed an average probability of correct prediction between 0.3224 and 0.3609, an average probability of the alternative not chosen between 0.2791 and 0.2959. The average probability of correct prediction was 2.65% to 8.18% higher than the average probability of the alternative not chosen, what is least desired. There was little likelihood that the average probability of correct prediction would be higher than 50% since the airline market in the region is very much competitive and all airlines have desired attributes from the passenger point of view. Since in most cases there are 3 or 4 airlines competing, an acceptable value should be higher than 33% or 25% respectively, what was found in fact.

The extremes of associability with the dependent variable were the model with the best associability (LFARE – lowest fare) and the model with the worst associability (IDF – indirect flight frequency in the travel day). Indeed the lowest fare is an essential tool the airlines use to attract the passengers, whereas indirect flights are poorly regarded by passengers.

Regarding the variables associated with the flight cost, the lowest fare (LFARE) was the most significant one, possibly because it means saving money to a higher extent than the highest fare, and this does not appeal to the passengers who are choosing an airline for their flights.

Among the variables of frequency, direct flight frequency in the travel day and period (DDPF) showed the best explanatory power on airline choice in the São Paulo Metropolitan Area. From the point of view of a connection or a stop on the way to the destination, the supply of direct flights better explained airline choice. It is evident that delays produced by a connection or a stop are undesired due to the loss of time, since rapidity is the main advantage of choosing the air mode of travel. Since total frequency of flights enhances the number of direct flights, it occupied second place in the ranking, better explaining airline choice than the variables of purely indirect flight frequency.

Among the variables of direct flight frequency, ranging from the one which best explains the airline choice to the one that has the lowest explanatory power, it was found: frequency in the travel day and period (DDPF), frequency in the travel period in all days of the week when the passenger traveled (DPF), frequency irrespective of day and period (DWF) and frequency in the travel day (DDF), respectively. The difference among the quality of the adjustment found was not significant, albeit perceivable. It was found that passengers are more prone to shift their departure date than their departure period of the day. Moreover, the departure period of the day (represented by DPF) was more significant than the day of departure itself (represented by DDF).

A possible explanation for the better adjustment of frequency of the departure period in all days of the week when the passenger traveled (DPF) in comparison to frequency in the travel day (DDF) is that passengers may show availability along the week to make their trips, but appointments with which they fulfill their schedule along the day may be regarded as a priority. For instance, on the one hand consider a businessman that must depart in the early morning from São Paulo to participate at a meeting at 10:00 a.m. in Belo Horizonte. On one hand there are plural options of days along the week when this meeting could be held; on the other hand, there is only one option of the period of time during the day which the meeting could be held. As another example, consider a worker living in São Paulo that decides to spend one week in the seaside of Rio de Janeiro beaches. To vary the period of departure along the day may mean poor scheduling of his trip, whereas it would not differ much if his trip were scheduled in the first or in the second week of his one month vacation.

Among the variables of indirect flight frequency, ranging from the one which best explains airline choice to the one that has the lowest explanatory power, it was found: frequency in the travel period in all days of the week when the passenger traveled (IPF), frequency in the travel day and period (IDPF), frequency irrespective of day and period (IWF) and frequency in the travel day (IDF), respectively. In this group the difference was not that big. On the other hand, the frequency along the week keeping the period of departure was more significant probably because the activities to be held at destination are scheduled in certain periods of the day.

Among the variables of total flight frequency, ranging from the one which best explains airline choice to the one that has the lowest explanatory power, it was found: frequency in the travel period in all days of the week when the passenger traveled (TPF), frequency in the travel day and period (TDPF), frequency irrespective of day and period (TWF), and frequency in the travel day (TDF), respectively. In this group the difference was not significant. Once again the frequency along the week of the period of departure was more important probably as a result of activities to be accomplished at destination occurring in certain periods of the day.

Finally, the airline age (AGE) was the second variable best explaining airline choice process, only losing to the lowest fare (LFARE). The airline age represents the length of time of the airline's presence in the market, the result of airline marketing strategies and the perseverance of a company which may have faced difficulties but succeeded in staying longer and is probably well-known by the majority of the nationals who usually rely on air transportation to develop their activities.

Models containing three explanatory variables

Among the models with three variables, 24 models have been tested, using all combinations of three variables among the 15 variables selected in the previous section, paying attention not to include in the same model variables of the same type. Therefore, for instance, total weekly flight frequency irrespective of day and period (TWF) and direct weekly flight frequency irrespective of day and period (DWF) were not tested in the same model specification because both of them are variables of the same flight frequency.

The models considering three variables enable the evaluation of the tradeoffs passengers face between the best choice variables of their airline selection. The best model for the entire sample considered the lowest fare (LFARE), the direct flight frequency in the day and period of departure (DDPF) and the airline age (AGE). This model was selected for further analysis of passenger market segments. The result of its calibration is shown in Table 4.

Table 4. Calibrations of the best model across market segments

Market segments	Lowest Fare	Direct Flight Frequency in the Travel Day and Period	Airline Age
Entire Sample	0.0067	0.1103	0.0227
Passengers departing from GRU	- 0.0006	0.5110	0.0211
Passengers departing from CGH	0.0072	0.0664	0.0228
Junior passengers	0.0053	0.1483	0.0207
Senior passengers	0.0084	0.0648	0.0253

It was also verified that in the models with three variables the signals of FARE were positive, and the explanation for this outcome is that passengers failed to choose the low-cost/low fare airline which was offering few flights but exhibited a great potential for expansion, what is now verified at the time of this publication, three years after the survey. Besides, the signals of the variables of indirect flight frequency were unexpectedly negative in the models where they appeared with LFARE. This did not pose a problem because these models did not present the highest average probability of the chosen alternative. Lastly, the signal of AGE was positive as expected in all the models of three variables, possibly as a result of a passenger preference for airlines longer in the market.

The t-Student statistics (whose presentation was omitted) were satisfactory in the case of the models containing variables of direct flight

frequency, presenting a modulus higher than 2. However, the models considering variables of indirect and total flight frequency, when associated with LFARE and AGE, produced t-Student statistics lower than 2 in modulus. Likewise, the p-value was somewhat high for the variables of frequency pertaining to these models. Therefore these models were regarded of lower reliability. The calibration of the models with three variables revealed an average probability of correct prediction between 0.3577 and 0.3899, an average probability of the alternative not chosen between 0.2664 and 0.2805. The average probability of correct prediction was 7.72% to 12.35% higher than the average probability of the alternative not chosen, what meant a reasonable improvement compared to the models of one variable. Once again, there was little likelihood that the average probability of correct prediction would be higher than 50%, even in a model of three variables, since the airlines face tight competition, each one specializing in one asset, be it the flight cost, the flight frequency or the airline performance. Since in most cases there are 3 or 4 airlines competing, then an acceptable value should be higher than 33% or 25% respectively, what happened in fact.

The best associability with the dependent variable was the model considering lowest fare (LFARE), frequency in the travel day and period (DDPF) and airline age (AGE). While the least associability with the dependent variable was the model considering highest fare (HFARE), frequency in the travel day (DDF) and airline age (AGE). Indeed having the lowest fare is an essential tool the airlines use to attract passengers, as opposed to the highest one. What was unexpected was that the worst model did not contain a variable of indirect flight frequency.

To analyze the tradeoffs between the variables of the best model, it is verified that the coefficient of direct flight frequency in the travel day and period (DDPF) is 16.46 times higher in modulus than that of LFARE. Therefore, through this model it is inferred that passengers pay R\$ 16.46 to bear the absence of each direct flight in the travel day and period to the desired destination supplied by this airline. This result is counterintuitive, albeit explained by the fact that the airline offering cheaper tickets had few flights so few passengers could actually choose this airline.

Moreover, the coefficient of AGE is 3.39 times greater in modulus than LFARE. Therefore, through this model it is inferred that passengers agree to pay R\$ 3.39 more for the travel ticket for each year the chosen airline is younger. This result is also counterintuitive, albeit explained by the fact that the airline offering cheaper tickets was the youngest (a new entrant) and few passengers actually chose this airline.

Last but not least, the coefficient of direct flight frequency in the travel day and period (DDPF) is 4.86 times higher in modulus than that of AGE. Therefore, through this model it is inferred that the chosen airline may be five years younger than the passenger desires in exchange for each direct

flight in the travel day and period to the desired destination supplied by this airline. This result is expected and intuitive.

Analysis of models in terms of passenger market segments

Having found the model with higher probability of the chosen alternative, which considered the variables the lowest fare (FARE), direct flight frequency in the day and period of departure (DDPF) and airline age (AGE), the passengers were segmented by airport of departure and passenger age, as Table 4 shows.

The signal of the coefficient of LFARE is negative, as expected for passengers departing from GRU (the expected passenger behavior is to select the airline with lower fares). Moreno and Muller (2004) showed that airport choice performed by passengers from GRU is not well explained by access time savings, so what really counts for these passengers is saving money with air fares. Moreover, the low-cost/low fare airline was not operating in GRU at the time of the interview with passengers; therefore, the fares offered to passengers in GRU were rather similar across airlines. This is interesting because at CGH there was the low-cost/low fare airline, but it offered few flights, so there was the possibility of flying this airline, but few passengers could do this in fact. On the other hand, passengers departing from CGH may also have been somehow careless about saving money with air tickets, probably because they are more worried about choosing the closer airport and end up choosing more expensive airlines.

Passengers from GRU place eight times more importance on direct flight frequency in the day and period of departure than those from CGH. This fact shows that passengers from GRU care very much about airline level of service and are aware of airline competition. However, Moreno and Muller (2004) showed that passengers probably do not consult flights from a competing airport, since their airport choice is not based on the rationality of an increase of flight supply.

Following on, passengers from CGH place more importance on airline age. CGH is also the older airport in the region, now aged more than 65 years old, whereas GRU is only 20 years old. Airline age is the result of succeeding in the market for several decades, and this is more promptly recognized by passengers departing from the oldest airport.

Senior passengers (over 36 years of age) are more careless about ticket price but place more importance on airline age. Senior passengers may have started flying late in life, so they are less concerned about prices of air tickets, but as a result of having lived longer, they may be more aware of airline marketing and announcing efforts than junior passengers (up to 36 years of age), so they are more worried about the variable of airline performance.

Finally, junior passengers place more importance on direct flight frequency in the day and period of departure. One easy way of consulting flight frequency is through the Internet, which appeals more to younger travelers, what can explain this result.

CONCLUSIONS AND RECOMMENDATIONS

Aiming at analyzing airline choice carried out by passengers departing the São Paulo Metropolitan Area, conditional LOGIT model was used as a modeling tool. Decision functions for each passenger were built, one for the chosen airline and one for each airline not chosen. Several specifications for the decision function were tested. These specifications enhanced independent variables pertaining to 3 groups: (a) variables related to flight cost; (b) variables accounting for flight frequency available at the analyzed airports; and (c) one variable associated with airline performance. The decision functions were built considering one or three of the variables described above, taking care not to mix variables of the same group in one model. The specification that produced the model most adjusted to the data (evaluated in terms of the highest average probability of the correct prediction) enhanced the following variables: lowest fare; direct flight frequency in the day and period of departure; and airline age.

Using the variables obtained from the best model, airline choice was analyzed segmenting the passenger market by departure airport and passenger age. From the analysis of the results achieved with the data collected for this work and for the region treated in this paper, it is possible to affirm the following:

1. The lowest fare is the factor that can best explain airline choice, despite its positive signal.
2. The variables of direct flight frequency exhibit better explanatory power on airline choice than variables of total flight frequency and the variables of total flight frequency exhibit better explanatory power on airline choice than variables of indirect flight frequency.
3. Airline age is the second best factor explaining airline choice.
4. Airline choice made by passengers departing from GRU is well explained by money savings, as opposed to airline choice made by passengers from CGH, which is not.
5. Airline age is more important for senior passengers, whereas the direct flight frequency in the day and period of departure is more important for junior passengers.

The recommendations are addressed to each group connected directly or indirectly with air transport activity. These recommendations were made up from this work, being restricted to its characteristics, such as seasonality of

the interviews along the year, the existing politic and economic scenario, delimitation of the trip origin region, studied destinations, model specifications and variables employed in the modeling. It is recognized that to put into practice any of these recommendations, caution is necessary as is validation of the conclusions of this work through periodic evaluations (studies) of airline choice in the São Paulo Metropolitan Area.

Since the airline choice at GRU is the result of money savings while at CGH it is not, this paper highlights that airline managers should have implemented a policy of higher fares in CGH and lower fares in GRU at the time of the interviews. Moreover, passengers departing from CGH and senior passengers should be focused by airline marketing strategies, as they are the market segments that most recognize airlines' efforts to stay longer in the market and airline performance, denoted by the variable airline age. Finally, airport managers should encourage airlines to schedule regular flights in the passengers' preferred day and period of departure, because at both airports airline choice is the result of an increase in the airline's level of service, denoted by the variable flight frequency.

Three alternatives are proposed to extend the research on airline choice in the São Paulo Metropolitan Area: (a) analyzing the importance of the variables pertaining to the best model across market segments according to other relevant criteria; (b) exploring other variables that may influence airline choice, such as the overall image of each airline according to passengers' opinions and a variable accounting for the flight schedule; and (c) carrying out airline choice research for international flights departing from GRU.

REFERENCES

- Armstrong, P., Garrido R., & Ortúzar, J. D. (2001). Confidence intervals to bound the value of time. *Transportation Research E*, 37(2-3), 143-161.
- Econometric Software, Inc. (2002). NLOGIT 3.0: Reference Guide. Plainview, USA.
- Hensher, D. A. (2001). The sensitivity of the valuation of travel time savings to the specification of unobserved effects. *Transportation Research E*, 37(2-3), 129-142.
- Kanafani, A. (1983). *Transportation demand analysis*. New York: McGraw-Hill.
- Koppelman, F. S., & Chu, C. (1985). Effect of sample size on disaggregate choice model estimation and prediction. *Transportation Research Record*, 944, 60-69.
- Lijesen, M. G. (2006). A mixed logit based valuation of frequency in civil aviation from SP-data. *Transportation Research E*, 42(2), 82-94.
- Mason, K. J. (2001). Future trends in business travel decision making. Air Transportation Research Society 2001 Conference (ATRS), Jeju, South Korea.

- Moreno, M. B., & Muller, C. (2003). Airport choice in São Paulo Metropolitan Area: An application of the conditional LOGIT model. 7th Air Transportation Research Society Conference (ATRS), Toulouse, France.
- Moreno, M. B., & Muller, C. (2004). Airport choice in São Paulo Metropolitan Area: An analysis based on passenger market segmentation. 8th Air Transportation Research Society Conference (ATRS), Istanbul, Turkey.
- Panrotas. (2002a). Tarifário Panrotas – Doméstico. In: Guia de horários e tarifas aéreas nacionais e internacionais, February, pp. B-75-B-119.
- Panrotas. (2002b). Tarifário Panrotas – Doméstico. In: Guia de horários e tarifas aéreas nacionais e internacionais, March, pp. B-75-B-119.
- Pels, E., Nijkamp, P., & Ritveld, P. (1998). Airport choice in a multiple airport region: An empirical analysis for the San Francisco Bay Area. Discussion Paper 98-041/3. Amsterdam, The Netherlands: Tinbergen Institute.
- Prousaloglou, K., & Koppelman, F. (1995). Air carrier demand: An analysis of market share determinants. *Transportation*, 22(4), 371-388.
- Suzuki, Y. (2000). The relationship between on-time performance and airline market share: A new approach. *Transportation Research E*, 36(2), 139-154.
- Suzuki, Y., & Walter, C. K. (2001). Potential cost savings of frequent flyer miles for business travel. *Transportation Research E*, 37(6), 397-410.
- Turner, S. (2003). Comparison of passenger profiles and selection criteria: A study of London–Amsterdam passengers. ATRS Conference, Toulouse, France.
- Windle, R., & Dresner, M. (1995). Airport choice in multiple-airport regions. *Journal of Transportation Engineering*, 121(4), 332-337.
- Yai, T.; Takada, K., & Okamoto, N. (1997). Analysis of international air passenger characteristics in the Asian region. The Eastern Asia Society for Transportation Studies, 2nd Conference, Seoul.
- Yoo, K. E., & Lee, Y. I. (2002). Air passengers' airline choice behavior for international flight. ATRS Conference, Seattle, WA, USA.