

AIRPORT CHOICE IN SAO PAULO METROPOLITAN AREA: AN APPLICATION OF THE CONDITIONAL LOGIT MODEL

Marcelo Baena Moreno

Energyarc Industrial Limitada
Avenida Regente Feijo, 813 – Agua Rasa – Sao Paulo – SP – Brazil – CEP: 03342-000
Phone: (55-11) 6128-5333
caixamarcelo@ig.com.br

Carlos Muller

ITA – Instituto Tecnológico de Aeronáutica
Praça Marechal Eduardo Gomes, 50 – Vila das Acacias – Sao Jose dos Campos – SP –
Brazil – CEP: 12228-900
Phone: (55-12) 3947-5738
muller@infra.ita.cta.br

Abstract

Using the conditional LOGIT model, this paper addresses the airport choice in the Sao Paulo Metropolitan Area. In this region, Guarulhos International Airport (GRU) and Congonhas Airport (CGH) compete for passengers flying to several domestic destinations. The airport choice is believed to be a result of the tradeoff passengers perform considering airport access characteristics, airline level of service characteristics and passenger experience with the analyzed airports. It was found that access time to the airports better explain the airport choice than access distance, whereas direct flight frequencies gives better explanation to the airport choice than the indirect (connections and stops) and total (direct plus indirect) flight frequencies. Out of 15 tested variables, passenger experience with the analyzed airports was the variable that best explained the airport choice in the region. Model specifications considering 1, 2 or 3 variables were tested. The model specification most adjusted to the observed data considered access time, direct flight frequencies in the travel period (morning or afternoon peak) and passenger experience with the analyzed airports. The influence of these variables was therefore analyzed across market segments according to departure airport and flight duration criteria. The choice of GRU (located neighboring Sao Paulo city) is not well explained by the rationality of access time economy and the increase of the supply of direct flight frequencies, while the choice of CGH (located inside Sao Paulo city) is. Access time was found to be more important to passengers flying shorter distances while direct flight frequencies in the travel period were more significant to those flying longer distances.

Keywords: Airport choice, Multiple airport region, Conditional LOGIT model, Access time, Flight frequencies, Passenger experience with the analyzed airports, Transportation planning

1. Introduction

The air transportation long term rising trend increases the regions worldwide served by more than one airport: the so-called "multiple airport regions". The study of the tradeoffs passengers face when flying out of these regions has therefore become a relevant subject in transportation planning.

In the multiple airport regions, the criteria used by passengers when they make their choice of departure airport constitute an important issue to several professionals, as follows:

1) The Air Transportation Authority: Because its mission is to study, orient, plan, control and motivate the Public and Private Civil Aviation activities, the segmentation of the market of passengers according to flight duration enables this organization to evaluate the importance of access time to the airport placed by the passengers and therefore enables scheduling of flights in the regional airports according to flight duration.

2) The Airport Managers: Searching for anticipated operational efficiency (such as a more consolidated terminal and high levels of passenger occupancy), they are worried with the capital destination to execute improvements rated as priorities by the air passengers, in special those that define the passenger choice of airport.

3) The Airlines: Aiming at a maximum seat factor at their aircrafts, the airport choice research enables the correct capital destination to schedule regular flights (aircraft allocation), distributing the flights among the airports in the multiple airport region according to the preferences pointed out by the passengers.

4) Urban Transportation Planners: These professionals are interested in becoming aware of the passengers' preferences for airports from the point of view of accomplishing public constructions that include improvement of accessibility to the airports, by means of railways or motorways, from the centers where the air transport demand originates.

The overwhelming majority of researchers on airport choice found two kinds of variables satisfactorily explaining the airport choice in a multiple airport region: one variable related to the accessibility to the competing airports and another variable accounting for the air transport level of service provided by these terminals.

This paper aims to determine which variables from a set of candidates have the best explanatory power on the airport choice made by the passengers whose travel starts in the Sao Paulo Metropolitan Area, as well as providing an analysis of the importance placed by the passengers on these variables when passengers are grouped by chosen airport and flight range.

The region of study is well served by two airports: Sao Paulo International Airport (GRU) and Congonhas Airport (CGH). They stand out in the Brazilian air transport system

in terms of the volume of embarked and disembarked passengers and compete for the air passengers for numerous domestic destinations.

This paper extends the research on airport choice by presenting and discussing results achieved in the analysis of airport choice in some multiple airport regions worldwide and bringing about results for the Sao Paulo Metropolitan Area.

2.Literature

With a LOGIT model, Skinner (1976) studied the airport choice in Baltimore-Washington. Building utility functions combining airline level of service measures and ground accessibility measures, the signals found for the variables were those expected. Skinner (1976) concluded that: a)The preferred level of service measure was the number of flight frequencies whereas the access utility (a combination of cost and time) was the best measure of ground accessibility; b)Non-business passengers are more worried about accessibility and/or less worried about the supply of flight frequencies; c)To shift the passengers' airport choice, an airport would have to supply an extremely higher amount of flight frequencies in relation to the competing airport; d)A ground accessibility improvement would be a positive policy to shift the passengers' airport choice if it were directed to the entire urban area.

Augustinus and Demakopoulos (1978) analyzed the airport choice in New York-New Jersey. To model the airport choice, they used a model that associates the probability that a passenger chooses an airport to the ratio between: a)The ratio of the costs associated to the cheapest airport over those associated to the chosen airport; b)The sum of the ratios of the costs of the cheapest airport over those of the alternative airports. These quotients a) and b) are previously powered by an exponent α that aims to capture the importance placed by passengers regarding cost differences between airports. In this model, the cost variable enhanced time cost and money cost. Augustinus and Demakopoulos (1978) concluded that access and factors related to convenience are more significant to passengers in short-haul flights, probably because they account for a higher percentage in the door-to-door travel time.

Harvey (1987) studied the airport choice in San Francisco Bay Area, using LOGIT model. He used as modeling variables: access time to the airport, absolute and relative (without connections) flight frequencies. Models exhibiting different formats were built, taking the variables either in their direct value or belonging to a parabolic function, the latter aimed at denoting the decreasing character of the rise in importance of the variable, whenever it was suitable. Harvey (1987) concluded that: a)From a certain supply of flight frequencies, an increase in this supply doesn't seem to make an airport more attractive; b)Direct flights are better regarded than connecting ones; c)the marginal disutility of the access time appears to decrease towards a rise in total time (inferred as total time of access); d)The importance passengers place on access time seems to vary across flight ranges; e)The airport choice seems to be dissociated from the access mode choice; f)For non-business passengers, both flight frequencies and access time are less important.

Ashford e Benchemam (1987) performed the analysis of the airport choice in Central England, using LOGIT model. The selected variables were: 1) access time from the origin of the travel to each airport; 2) daily flights to the chosen destination; 3) economy fare. The market of passengers was segmented in domestic, business international, leisure international and inclusive tours international. The variable "fare" was excluded from the analysis for business and inclusive tours because its parameter was positive. For business passengers and inclusive tours, the access time was found to be a dominant factor in comparison to flight frequencies, whereas the fare was found to be a dominant factor for domestic and leisure passengers. It was concluded that non-business passengers place more importance on the access and less importance on the flight frequencies in comparison with business passengers.

Innes and Doucet (1990) studied the airport choice in the north of New Brunswick province, Canada, using binary LOGIT model through LOGIST procedure. Variables associated with the air transport level of service and access distance were tested. However, since the signal of the parameter of the access distance variables was positive, these variables were excluded from the modeling. Among the models considering only level of service variables, the variable "aircraft type" was more important to residents, who traveled significant distances in search of a jet service. Other relevant variables were "difference between flight durations" and whether direct flights were offered to the passenger's destination or not. The passengers rejected commuter service, choosing airports with direct flights even when located farther, to an extent that this service was discontinued after this study.

Windle and Dresner (1995) studied the airport choice in Baltimore-Washington D.C., using multinomial LOGIT model, inferring that: 1)The access time was more important to business passengers; 2)The weekly flight frequencies are better regarded by the non-residents in the region; 3)The weekly flight frequencies are more important to business passengers; 4)The inclusion of a variable accounting for the experience with DCA (National) and IAD (Dulles) airports decreases the probability of choosing BWI (Baltimore-Washington Intl' airport) and the importance of the access time; 5)Duplicating the previous result, but only for passengers that accessed the airport with motor vehicles, it was noted more importance placed on access time; 6)Passengers originating from zones with fairly equivalent access time place more importance on weekly flight frequencies and less importance on access time, not depending on the market segment; 7)The passengers' experience with an airport is generally an important determinant of the airport choice.

Pels, Nijkamp and Rietveld (1998) studied the conjoint choice of airport and airline in the San Francisco Bay Area, using nested LOGIT model, building two situations of sequential choice: 1)First airport choice and then airline choice; and 2)First airline choice, followed by airport choice. Pels, Nijkamp and Rietveld (1998) didn't find 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 the airport choice happens first, and then the airline choice, not being simultaneous choices.

Researches treating the same region in different years showed distinct passengers' appraisal on accessibility and level of service measures, in terms of market segments grouped by travel purpose. As a result, the evolution of passenger behavior across time indicates the demand for seasonal researches on airport choice in the same multiple airport region.

3. Model

The LOGIT model has been the most widely used model 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 (airport) whose utility is the maximum among the available choices. This choice behavior can be expressed mathematically by the following inequation:

$$U_{mi} \geq U_{ni} \text{ for all } n, 1 \leq n \leq z \quad (1)$$

Where: U_{mi} is the utility that passenger i obtains by choosing airport m , U_{ni} is the utility that passenger i obtains by choosing airport n , z is the number of airports (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:

$$U_{mi} = D_{mi} + R_{mi} \text{ for all } m, 1 \leq m \leq z \quad (2)$$

Where: U_{mi} is the utility that passenger i obtains by choosing airport m , D_{mi} is the deterministic part of the utility function for alternative m chosen by passenger i , R_{mi} is the random part of the utility function for alternative m chosen by passenger i , z is the amount of choices considered available for passenger i .

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 is given by:

$$p_m (D_1, \dots, D_m, \dots, D_z) = \frac{e^{\alpha \cdot D_m}}{\sum_{z=1}^z e^{\alpha \cdot D_z}} \quad (3)$$

Where: p_m is the choice probability of the alternative m (each alternative is an airport in this paper), among the z alternatives (airports); D_m is the deterministic part of the utility function of alternative m (airport); α is the parameter associated with the deterministic part of the utility function.

While at an individual level the formulation predicts the probability that a passenger chooses the airport in question, at an aggregate level it predicts the share of the considered airport.

To accomplish the estimation of parameter α , NLOGIT program version 3.0 was used, available in the LIMDEP program version 8.0 produced by the company Software Econometric, Inc. The reference guide of this software classifies the multinomial LOGIT models:

1) Models whose variable values are input the same across all alternatives for the same observation (passenger), as they are individual characteristics;

2) Models whose variable values are attributes of the alternatives (perceived by passengers), and variables whose value remain constant across alternatives (for the same passenger) are also allowed.

The latter is the model that this paper employed, also known as conditional LOGIT model, which estimates variable parameters using the Maximum Likelihood Method. For the iterations, Newton Method was used because it produced quick convergence for all calibrated models.

The goodness-of-fit measure generated by the Conditional LOGIT model of this software is called R^2 ($0 \leq R^2 \leq 1$). However, the percentage of passengers whose actual choice is similar to the one the model predicts is given by the fraction of concordant pairs (and not by the value of R^2 , as the users of linear regression may think).

4. Case Study

Although the Sao Paulo Metropolitan groups several towns, 7 of them (Sao Paulo, Guarulhos, Santo Andre, Sao Bernardo do Campo, Sao Caetano do Sul, Diadema e Osasco) have been chosen to represent the trip origin in this region, because of two reasons:

a) They represent 79% of the electric power consumption in the region;

b)The calculation of the access time from the other towns wasn't likely to lead to sound values.

The alternative airports whose choice was analyzed were Congonhas Airport (CGH, located in Sao Paulo city) and Guarulhos Airport (GRU, located in Guarulhos city, neighboring Sao Paulo). The criteria for destination selection was that there must have been departures to these destinations from both airports and that the annual volume of passengers must have surpassed 100,000 passengers. This volume, as Windle and Dresner (1995) tell, avoids small sample bias, that usually represent less popular destinations. These two requisites were evaluated through the last statistical report of the DAC (Department of Civil Aviation) available to the date of data collection, the report of the year 2000.

Therefore 21 airport destinations (19 cities) were studied in this paper: 1)BPS (Porto Seguro); 2)BSB (Brasilia); 3)CGR (Campo Grande); 4)CNF (Belo Horizonte); 5)CWB (Curitiba); 6)FLN (Florianopolis); 7)FOR (Fortaleza); 8)GIG (Rio de Janeiro); 9)GYN (Goiania); 10)IGU (Foz do Iguacu); 11)JOI (Joinville); 12)LDB (Londrina); 13)NVT (Navegantes); 14)PLU (Belo Horizonte); 15)POA (Porto Alegre); 16)RAO (Ribeirao Preto); 17)REC (Recife); 18)SDU (Rio de Janeiro); 19)SSA (Salvador); 20)UDI (Uberlandia); 21)VIX (Vitoria);

The passenger profile was obtained by revealed preference interviews at the airports, conducted during the weekdays of two consecutive weeks (February 8th to March 1st, 2002, during the peak hours of access to airports, i.e., from 7 AM to 10 AM (morning peak) and from 5 PM to 8 PM (afternoon peak).

These periods were chosen because the average vehicle speeds in Sao Paulo city have been measured during these periods by CET (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 minimization of time and cost of data collection, compilation and analysis, 1923 passengers were interviewed, being 897 at GRU and 1026 at CGH. This amount of observed data has been considered satisfactory taking into account the paper of Koppelman and Chu (1985) who calculated the amount of observations required for relatively simple disaggregate choice models.

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

5. Airport choice: explanatory variables

Three types of variables were chosen to be tested: those associated with the ground accessibility to the airports, those related to the airlines' level of service in the airports and the last one is associated with the passenger experience with the airports in the studied region.

Table 1: Results of the interview with passengers

Sample segmentation criteria	Passenger market segments
Departing period	853 departing in the morning peak and 1070 in the afternoon peak
Departing airport	897 departing from GRU and 1026 from CGH
Travel purpose	1400 business and 523 non-business
Place of residence	937 residents and 986 visitors
Age	975 under 37 years old and 948 over 37
Household monthly income	1024 under R\$ 6,000.00 monthly wage and 899 over R\$ 6,000.00
Passenger experience with the airports in the region	946 under 7 flights during the year 2001 and 977 over 7
Access mode	127 accessing by bus and 1745 accessing by taxi or car (either hitchhiking or driving)
Declared reasons for choosing the departing airport	290 didn't know that there were flights at the competing airport to the desired destination; 287 choice was made by a third party; 902 choice by proximity and 460 choice by the supply of more flight frequencies to the desired destination.
Flight range	1116 in flights with flying time up to one hour and 807 in longer flights
Type of airline chosen	1883 using regular cost airlines and 40 using low cost airlines

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, e.g., the coefficients of the variables were the same for both alternative airports, GRU and CGH.

Each passenger i , $1 \leq i \leq 1923$, has been represented by two generic decision functions, being the first always for GRU and the second for CGH, whichever the departing airport, as follows:

$$CHOICE_{i\ GRU} = \alpha_1 \cdot ACCESS_{i\ GRU} + \alpha_2 \cdot FREQUENCY_{i\ GRU} + \alpha_3 \cdot EXPERIENCE_{i\ GRU} \quad (4)$$

$$CHOICE_{i\ CGH} = \alpha_1 \cdot ACCESS_{i\ CGH} + \alpha_2 \cdot FREQUENCY_{i\ CGH} + \alpha_3 \cdot EXPERIENCE_{i\ CGH} \quad (5)$$

Below the choice probability of each airport by each passenger is shown:

$$P_{i\ GRU} = \frac{\exp(CHOICE_{i\ GRU})}{[\exp(CHOICE_{i\ GRU}) + \exp(CHOICE_{i\ CGH})]} \quad (6)$$

$$p_{i \text{ CGH}} = \frac{\exp(\text{CHOICE}_{i \text{ CGH}})}{[\exp(\text{CHOICE}_{i \text{ GRU}}) + \exp(\text{CHOICE}_{i \text{ CGH}})]} \quad (7)$$

Where: ACCESS is a variable associated with the access to the airports; FREQUENCY is a variable associated with the airlines' level of service; EXPERIENCE is a variable related to the passenger's experience with the airports in the region; $p_{i \text{ GRU}}$ is the probability that passenger i chooses GRU; $p_{i \text{ CGH}}$ is the probability that passenger i chooses CGH; CHOICE is a decision function of airport choice; α_k is the parameter (coefficient) related to the variable k , being $k=1$ for ACCESS, $k=2$ for FREQUENCY and $k=3$ for EXPERIENCE.

5.1. Variables associated to the access to the airports (ACCESS)

Since 100% of the passengers faces the tradeoff between the access distances (AD) to the alternative airports, this variable was tested, being expressed in km.

The access time (AT) by auto mode (in this paper denoted by "general traffic") was the variable most widely used in the analyzed literature. Because 1745 passengers (91%) face the tradeoff between access times to the competing airports using general traffic when they make an airport choice, this variable was tested, its value being expressed in minutes.

To calculate the access time to each airport, the studied region was sectioned in 101 zones (95 in Sao Paulo city and the other 6 towns constituted, one by one, other 6 zones). The access distances from the centroids of the zones were measured considering the access by the existing streets, avenues and roads.

The average speeds along the main streets and avenues were extracted from the CET (Traffic Engineering Company) annual reports, years 2000 and 2001. These speeds were grouped by access mode: bus and general traffic, and they account for the period of the day: morning peak hours (7 AM to 10 AM) and afternoon peak hours (5 PM to 8 PM). Data from different years were used as a matter of availability and updating at the time of their collection. This didn't pose a problem because there were no significant modifications in the access to the airports from the first year considered until the period of interview with the passengers.

Regarding the streets and avenues whose speed was not available, average values of the measured ones were adopted, weighing by length of the street or avenue, and considering the period of the day. For highways, a value of 80 km/h was adopted, irrespective of the period of the day.

5.2. Variables associated the airlines' level of service in the airports (FREQUENCY)

Among the level of service variables, flight frequencies were the most commonly used in the analyzed literature, and a satisfactory explanatory power was verified, therefore 12 flight frequencies-related variables have been tested.

These variables were built in terms of the following criteria: 1)The existence of connections or stops (direct flights, indirect flights and the sum of the two); 2)The travel period (morning peak or afternoon peak); 3)The day of the week.

In terms of the second criterium, the passengers were interviewed at the moments prior to their departure, at times before their proceeding to the waiting room and sometimes after. The morning peak was considered from 7 AM to 10 AM and the afternoon peak from 5 PM to 8 PM. Albeit the data collection was after the access to the airport, it was admitted that close to the peak periods (no matter before or after them) the highway access speeds do not vary at a significant basis from the peak periods themselves.

The flight frequencies across periods of the day and across days of the two weeks when the interview took place were determined through the Internet websites of the airlines that offer regular flights and operate at the competing airports. Although the interviews had taken place during the weekdays, weekend flight frequencies were also accounted for since they increase the utility associated with the alternative airport where they are supplied.

For each of the built variables of frequency, its value was collected for the chosen airport and the airport not chosen. Below these variables are presented: 1)DDPF: Direct frequencies in the travel day and period; 2)DDF: Direct frequencies in the travel day; 3)DPF: Direct frequencies in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled; 4)DWF: Direct weekly frequencies irrespective of day and period; 5>IDPF: Indirect (with connections or stops) frequencies in the travel day and period; 6>IDF: Indirect frequencies in the travel day; 7)IPF: Indirect frequencies in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled; 8)IWF: Indirect weekly frequencies irrespective of day and period; 9)TDPF: Total (direct plus indirect) frequencies in the travel day and period; 10)TDF: Direct frequencies in the travel day; 11)TPF: Direct frequencies in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled; 12)TWF: Direct weekly frequencies irrespective of day and period.

Consider a passenger that departed to a given destination in the morning peak of February 19th, 2002. The variables of direct flight frequencies are shown in Table 2, and the variables of indirect and total flight frequencies were built in an analogous way:

Table 2: Variables of direct flight frequencies

	18	19	20	21	22	23	24	Total
	Mon	Tue	Wed	Thu	Fri	Sat	Sun	
Up to 7h								
7h-10h		DDPF						DPF
10h-17h								
17h-20h								
After 20h								
Total		DDF						DWF

Most flights to Rio de Janeiro central airport, Santos Dumont (SDU), depart from Sao Paulo central airport, Congonhas (CGH), whereas most flights to Rio de Janeiro international airport, Tom Jobim (GIG), far from the city center, depart from Sao Paulo international airport, Guarulhos (GRU), which neighbors Sao Paulo city.

In the same way, most flights to Belo Horizonte central airport, Pampulha (PLU), depart from Sao Paulo central airport, Congonhas (CGH), whereas most flights to Belo Horizonte international airport, Confins (CNF), far from the city center, depart from Sao Paulo international airport, Guarulhos (GRU), which neighbors Sao Paulo city.

For these reasons, a question on the reasons for airport choice in the interview with the passengers included whether the passenger chose the airport of departure because a considerable surplus of flight supply to the desired destination airport in case the passenger destination was also a multiple airport region (Rio de Janeiro and Belo Horizonte). In case the passenger answer was positive, frequencies to the other airport in the multiple airport region of destination were not included in the several variables of flight frequencies. For instance, a passenger flying to Rio de Janeiro and departing from CGH, if he declared his airport choice because he prefers to arrive at SDU, frequencies to GIG, the competing destination airport, were not considered no matter they were supplied at GRU or CGH.

5.3. Variable related to the passenger experience with airports (EXPERIENCE)

Windle e Dresner (1995) used this variable, showing that incorporating a variable that accounts for the passenger experience with an airport of departure in the studied region (in this case the number of times that the passenger departed from each airport in the region during the previous year), the importance placed on access time and weekly flight frequencies decreases, irrespective of the market segment (they segmented the market by travel purpose and place of residence).

This variable was tested in the present paper, being expressed by the number of domestic departures in the year 2001 from each airport, CGH and GRU. International flights (available at GRU) were not considered because CGH has only been serving domestic destinations lately. The fact that the traffic volumes considered at the studied airports are from year 2000, in function of availability and updating, doesn't spoil in any sense the data collection of this variable since the questions treated are from different nature.

5.4. Overall considerations on 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 2 or 3 variables). To begin with, 77 models were calibrated. Variables belonging to the same category didn't take part of the same model.

Table 3 exemplifies the decision functions of a fictitious passenger that flew from GRU to BSB, departing in the morning peak of February 25th, 2002, having left the zone

called "Penha" (in Sao Paulo city) in the day of his travel. Admit that this passenger departed twice from GRU and once from CGH during the year 2001. For one model built with three variables such as AT, DPF and EXP, equations (8) and (9) are obtained.

Table 3: Variables AT, DPF and EXP for a fictitious passenger

	AT (min)	DPF	EXP
GRU	29.9	21	2
CGH	53.2	49	1

$$1 = \frac{\exp(\alpha_1 \cdot 29.9 + \alpha_2 \cdot 21 + \alpha_3 \cdot 2)}{[\exp(\alpha_1 \cdot 29.9 + \alpha_2 \cdot 21 + \alpha_3 \cdot 2) + \exp(\alpha_1 \cdot 53.2 + \alpha_2 \cdot 49 + \alpha_3 \cdot 1)]} \quad (8)$$

$$0 = \frac{\exp(\alpha_1 \cdot 53.2 + \alpha_2 \cdot 49 + \alpha_3 \cdot 1)}{[\exp(\alpha_1 \cdot 29.9 + \alpha_2 \cdot 21 + \alpha_3 \cdot 2) + \exp(\alpha_1 \cdot 53.2 + \alpha_2 \cdot 49 + \alpha_3 \cdot 1)]} \quad (9)$$

6. Results

6.1. Models considering only 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 bring out the variable with best explanatory power on the airport choice in Sao Paulo Metropolitan Area. Table 4 presents the calibration results of these models.

The signals for the coefficients were those expected: negative for ACCESS and positive for FREQUENCY and EXPERIENCE. Indeed the increase on access distance and time are undesired, whereas the supply of flights and the experience a passenger has with an airport are desired and their increase propels the airport choice.

The t-Student statistics were satisfactory, presenting 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 R² values among 0.01041 and 0.17978. The associability extremes with the dependent variable were the model considering EXP (best associability) and the model considering IPF (worst associability). Since the fraction of concordant pairs in all the cases fell between 51% and 61% of the data, the adjustments were regarded as of medium quality.

IPF presented low explanatory power on the airport choice, probably because an indirect flight (enhancing stops or connections) is extremely undesired whenever direct flight frequencies are available to the chosen destination.

Table 4: Models with only 1 explanatory variable

Model	Variable	R ²	Coefficient	t-Student	p-value	Fraction of concordant pairs
1	AD	0.04941	-0.0185	-11.427	0	0.54
2	AT	0.05931	-0.0140	-12.210	0	0.54
3	DDPF	0.12613	+0.1951	16.123	0	0.59
4	DDF	0.11768	+0.0305	14.921	0	0.58
5	DPF	0.13269	+0.0336	16.593	0	0.59
6	DWF	0.12468	+0.0054	15.336	0	0.58
7	IDPF	0.01133	+0.0613	6.155	0	0.51
8	IDF	0.02146	+0.0146	7.945	0	0.52
9	IPF	0.01041	+0.0099	5.959	0	0.51
10	IWF	0.01878	+0.0024	7.525	0	0.52
11	TDPF	0.07890	+0.0883	13.809	0	0.55
12	TDF	0.08634	+0.0154	14.055	0	0.56
13	TPF	0.08235	+0.0154	14.048	0	0.56
14	TWF	0.09081	+0.0028	14.311	0	0.56
15	EXP	0.17978	+0.1023	14.780	0	0.61

Yet EXP was the variable exhibiting the best explanatory power on the airport choice in Sao Paulo Metropolitan Area, probably because the more a passenger travels, the more used he gets to the tradeoffs pertaining to an airport choice, as well as he becomes aware of these tradeoffs, creating a consistent structure of terminal choice and accumulating more experience of departure from one of the airports.

Among the variables of frequency, direct flight frequencies in the period of departure (DPF) showed best explanatory power on the airport choice in Sao Paulo Metropolitan Area. The ranking of the variables of frequency was notorious: the best adjustments fell among the direct frequencies, followed by the total frequencies and lastly were observed the indirect frequencies.

From the point of view of a connection or a stop on the way to the destination, the supply of direct flights was better explained the airport choice. It is evident that delays produced by a connection or a stop are undesired due to the loss of time, since the rapidity is the main advantage of choosing the air mode. Therefore, as the total frequencies enhance the direct ones, total frequencies occupied the second place in the ranking, better explaining the airport choice than the purely indirect flight frequencies.

Among the direct flight frequencies, ranging from the one which best explains the airport choice to the one that has the lowest explanatory power, it was found: DPF, DDPF, DWF and DDF. 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 and period of departure itself (represented by DDPF).

A possible explanation for the better adjustment of DPF in comparison to 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 Sao Paulo to participate at a meeting at 10 AM in Belo Horizonte. On the other hand there are plural options of days along the week when this meeting could be held. As another example consider a worker living in Sao Paulo that decides to spend one week in the seaside of Rio de Janeiro beaches. Vary the period of departure along the day may mean poor scheduling of his trip, whereas there wouldn't differ much if his trip were scheduled in the first or in the second week of his one month vacation.

Among the indirect frequencies, ranging from the one which best explains the airport choice to the one that has the lowest explanatory power, it was found: IDF, IWF, IDPF and IPF. Also in this group the difference was not that big. On the other hand the frequencies along the day were more significant probably because when the passenger faces scarcity of direct flight frequencies during the airport choice process, he becomes careless about shifting his flight schedule along the day since he will undoubtedly affront the vicissitudes of a delay caused by a connection or stop on his way to the destination.

Among the total frequencies, ranging from the one which best explains the airport choice to the one that has the lowest explanatory power, it was found: TWF, TDF, TPF e TDPF. Also in this group the difference was not that big. Maybe the weekly frequencies assumed the leadership in the airport choice process because of the importance of the multitude of options when total frequencies are analyzed, i.e., when the passenger doesn't have any guarantee that he will fly with or without a connection or stop.

Regarding the variables of ground accessibility, access time was most significant one, possibly because the passengers not only prefer the closest airport to the trip origin but also this proximity must be expressed much more in terms of time economy than in distance economy.

6.2. Models with two explanatory variables

From this category, 38 models were built, considering all combinations of two variables among those selected in the previous section of this paper, paying attention not to input variables of the same type in one model. The analysis of these models enables the evaluation of the tradeoffs passengers face between the best choice variables at their airport selection. Table 5 shows the calibration results of the models with higher R^2 .

In most models with two variables, the signals of the coefficients were those expected: negative for ACCESS and positive for FREQUENCY and EXPERIENCE. However, the signals of IDPF and IPF were negative in the models calibrated with AD and EXP. Since the modulus of the t-Student statistics was lower than 2 and the p-value of the variables of indirect flight frequencies was high in many cases, the inclusion of indirect frequencies showed itself undesirable to represent the airport choice in certain specifications of the decision function.

Table 5: Best models considering 2 variables (higher R^2)

Model	Variable 1	Variable 2	R^2	Coefficient 1 (t-Student)	Coefficient 2 (t-Student)	Fraction of concordant pairs
1	AT	DPF	0.15449	-0.0092 (-7.459)	+0.0301 (14.350)	0.60
2	AT	EXP	0.19490	-0.0079 (-6.249)	+0.0932 (13.454)	0.62
3	DPF	EXP	0.23410	+0.0240 (11.226)	+0.0820 (11.932)	0.65

The calibrations of the models considering two variables revealed R^2 values between 0.04942 and 0.23410. The associability extremes with the dependent variable were the model considering EXP and DPF (best associability) and the model considering AD and IWF (worst associability). Since the fraction of concordant pairs in all the cases fell between 54% and 65% of the data, the adjustments were regarded as medium quality.

As expected, a model considering one variable of indirect flight frequencies was the least adjusted to the data, since the passengers obviously prefer direct and total flight frequencies, as exposed before. Yet the access distance is less preferred than the access time, since the time economy is intrinsic to the air transportation as a travel mode, then the time economy during the access to the terminal is also more searched than a shorter distance.

The experience once again appeared in the best model. A possible explanation is that the experience with airport results from a detailed trip choice process, then the airport choice for the trip that took place during the day of the interviews should be connected to the passenger experience with that terminal in previous occasions. Besides experience, DPF reflected that passengers probably exercise their right to consult each airport's flight schedules well in advance of the ticket emission, thus preferring direct flights that bring about higher utility than the delays of indirect flights

The models with two variables have been classified in three groups: a) those considering variables of access and frequency; b) those with variables of access and experience and c) those with variables of frequency and experience. Among these three configurations, the last one provided the most adjusted model to the database ($R^2 = 0.23410$), involving the variables DPF and EXP. Thus once again experience was found to contribute to the formation of a consistent airport choice structure. Moreover, it was noticed again that passengers prefer the direct flight frequencies.

As well as observed in the models with one variable, the models with two variables considering one variable of indirect flight frequencies presented the lowest R^2 value. Once again was shown the fact that connections and stops on the way to the destination are undesired in comparison with total and direct flight frequencies.

6.3. Models of 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, TWF and DWF were not tested in the same model specification because both of them are variables of flight frequencies.

In the same way as the models with two variables, the models considering three variables enable the evaluation of the tradeoffs passengers face between the best choice variables at their airport selection. The best model considered access time to the airport, direct flight frequencies in the period of departure and passenger experience with analyzed airports. This model was selected for further analysis of passenger market segments. The result of its calibration is shown in table 6.

Tabela 6: Model with three explanatory variables that presented the highest R^2

Model	Variables			R^2	Coefficients (t-Student)			Fraction of concordant pairs
	1	2	3		1	2	3	
1	AT	DPF	EXP	0.24011	-0.0052 (-3.975)	+0.0226 (10.346)	+0.0775 (11.248)	0.65

It was also verified that in the models with three variables the signals of the coefficients were usually those expected: negative for access and positive for frequencies and experience. In the cases which this didn't occur, the modulus of the t-Student statistics was low, around 2, and the p-value of their variables differed from zero, indicating that the inclusion of the variables related to these coefficients are undesirable to represent the airport choice in the considered specifications of the decision function.

Because the fraction of concordant pairs in all the cases fell between 62% and 65% of the data, the adjustments were considered of medium quality.

The models with three variables can be classified by two basic criteria: a) by the variable of access and b) by the variable of flight frequencies. The variable of experience pertained to all the three-variable-models.

Grouping the models with three variables by the variable of access, two groups are formed: those which consider AT and those which consider AD. The quality measure indexes of the adjustment of the models with AD were lower than those which considered AT, probably because time economy during the access to the terminal is also more searched than a shorter distance.

Considering the frequency criteria, the models with three variables were divided into three groups: a) those which considered direct flight frequencies; b) those which considered indirect flight frequencies and c) those which considered total (direct plus indirect) flight frequencies. Once again the measures of goodness of fit were lower

whenever indirect flight frequencies were taken into account, whereas the measures of goodness of fit were higher in the case with direct flight frequencies.

The calibrations of the models with three variables revealed R^2 values among 0.18634 and 0.24011. In two extremes of associability with the dependent variable, the variable EXP has taken place, and EXP has already proved to play a significant role in the airport choice. What has differentiated these models was the presence of AT and DPF constituting the highest associability and AD and IDF constituting the lowest associability. Once again an indirect flight frequency (with connection or stop) has little influenced the airport choice in comparison to those direct and total. Moreover, the variable AD is less preferred than AT, as said previously, the time economy is intrinsic to the air transport mode choice, thus the time economy plays a more important role on the airport choice than economy in terms of distance.

To analyze the tradeoff between the variables of the best model, it is verified that the coefficient of DPF is 4.35 times higher in modulus than that of AT. Therefore through this model it is inferred that passengers put up with taking about further 4 minutes to one airport in exchange of each direct flight frequency in the period of departure this airport offers to the desired destination. For distant zones from one airport, where the access time exceeds one hour in certain periods of the day, 4 minutes further in the access wouldn't matter much, and also an increase of one direct flight frequency would be extremely beneficial in the case of a destination poorly supplied by direct flights. On the other hand for close zones to one airport, further 4 minutes in the access do bring a slight inconvenience, and also an increase of one direct flight frequency wouldn't mean a significant improvement in the case of the most popular destinations.

Moreover, the coefficient of EXP is 14.90 times greater in modulus than AT. Therefore through this model it is inferred that passengers put up with taking further 15 minutes to one airport for each experience of departure he had for domestic flights using this airport in the year before the interview took place.

Last but not least, the coefficient of EXP is 3.43 times higher in modulus that that of DPF. Therefore through this model it is inferred that for each experience with domestic flights a passenger had in the previous year with one airport, he puts up with the absence of supply of three flight frequencies to the desired destination in that airport.

A practical and interesting illustration of what this model reveals is the calculation of the percentage distribution between airports, of the passengers without any experience in the year 2001 with GRU and CGH airports. In this case, even considering a model of three variables, the airport choice is based on the tradeoff between only AT and DPF. Suppose for exemplifying reasons that these passengers leave zone 55 (Mooça in Sao Paulo), traveling in the morning peak of February 19th, 2002 to (destination BSB). Table 7 shows the variables related to their choice.

Table 7: Variables AT, DPF e EXP for fictitious passengers.

I	Airport	AT	DPF	EXP
1	GRU	32.4	21	0
2	CGH	31.9	42	0

$$\exp(U_1) = \exp(-0.0052 \cdot 32.4 + 0.0226 \cdot 21 + 0.0775 \cdot 0) = \exp(0.3061) = 1.3581$$

$$\exp(U_2) = \exp(-0.0052 \cdot 31.9 + 0.0226 \cdot 42 + 0.0775 \cdot 0) = \exp(0.7833) = 2.1887$$

$$\Sigma \exp(U_i) = 3.5468$$

$$p_1 = \exp(U_1) / \Sigma \exp(U_i) = 1.3581 / 3.5468 = 38\%$$

$$p_2 = \exp(U_2) / \Sigma \exp(U_i) = 2.1887 / 3.5468 = 62\%$$

6.4. Analysis of the variables of the chosen model in terms of passenger market segments

Having found the model with higher R^2 , which considered the variables access time, direct flight frequencies in the period of departure and experience, the interviewed passengers were segmented by airport of departure and flight range, as Table 8 shows.

Table 8: Calibrations of the model with greater R^2 across some market segments

Market segments	AT	DPF	EXP
Passengers departing from GRU	0.0242	-0.0064	0.0248
Passengers departing from CGH	-0.0752	0.0587	0.1324
Passengers in shorter-haul flights	-0.0184	0.0075	0.1042
Passengers in longer-haul flights	-0.0035	0.0444	0.0665

The signals of the coefficients of AT and DPF are not counterintuitive in the case of GRU (the expected passenger behavior is to select the closest airport to the trip origin and the airport with a superior supply of direct flight frequencies to the desired destination, as far as possible), while at CGH the signals were not only the expected ones but also their values were higher than those encountered for sample prior to segmentation.

Regarding the access, the positive signal of AT at GRU indicates that the passengers that have chosen GRU for their trips didn't show priority for the variable of access time, since GRU is the farthest of the two analyzed airports from the center that generates air transport demand in the Sao Paulo Metropolitan Area. This is verified by the random sample collected: among the zones enhancing 50 or more interviewees, only one zone is closer to GRU in terms of access time, whereas the 13 other zones with more than 50 interviewees are closer to CGH.

Meanwhile a possible explanation for the correct negative signal and high value of the coefficient of AT in CGH is that the center where air transport demand is generated lies closer to CGH than to GRU.

Regarding airlines' level of service at the competing airports, the high value of DPF in CGH in comparison to the sample as a whole is justified by the observation of only domestic flights departing from CGH, whereas GRU's runway is shared by domestic and in large scale international flights. In the case of some destinations like Rio de Janeiro and

Belo Horizonte, the choice of CGH as an airport of departure is by far related to the desire to land respectively at SDU and PLU, whereas the departures to their respective competing airports GIG and CNF, whose demands are comparatively smaller, are somewhat concentrated at GRU.

Finally the experience is a more important factor in CGH than in GRU. Very many CGH passengers have already structured their airport choice process, and repeat this option as far as possible (whenever there is flight supply to the desired destination).

The coefficient of the variable of access time was presented a higher modulus for passengers in short-haul flights. More importance on access time is expected for this market segment, which searches the closest airport, since access time means a larger part of the door-to-door time of the trip for those flying shorter distances.

Yet, the coefficient of DPF was higher for passengers in long-haul flights. It is expected that these passengers place more importance on airline level of service, since they are going to remain a longer time inside the aircraft. Besides, it is more common for long-haul flights to include a connection or stop (and thus they are more time-consuming) than for short-haul ones. Therefore long-haul flight passengers search more enthusiastically an airport that offers direct flight frequencies.

7. Conclusions and recommendations

Aiming at analyzing the airport choice made by the passengers in the Sao Paulo Metropolitan Area, conditional LOGIT model was used as a modeling tool.

Decision functions for each passenger were built, one for the chosen airport and another for the airport not chosen.

Several specifications for the decision function were tested. These specifications enhanced independent variables pertaining to 3 groups: a) variables related to the airport access; b) variables related to the airlines' level of service in the airports; c) one variable accounting for the passenger's experience with the airports in the region.

The decision functions were built considering one, two 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 higher R^2) enhanced the following variables: access time, direct flight frequencies in the period of departure and passenger's experience with the airports in the studied region.

Using the variables obtained from the best model, the airport choice was analyzed in terms of a passenger market segment by airport of departure and flight range.

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 that, generally:

1. The experience a passenger has with an airport of departure is the factor that can better explain the airport choice;
2. The access time to an airport better explains the airport choice than the access distance;
3. The direct flight frequencies exhibit better explanatory power on the airport choice than those total and the latter on those indirect;
4. The choice of GRU is not well explained by access time economy savings and by an increase in the supply of direct flight frequencies in the period of departure to the desired destination, contrarily to the choice of CGH, which is;
5. The access time is more important to those passengers in short-haul flights, whereas the supply of direct flight frequencies in the period of departure is more important to those passengers flying longer distances;

The recommendations are addressed to each group connected directly or indirectly with the 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, variables employed in the modeling, and so on. It is recognized that to put into practice any of these recommendations, it is necessary caution and validation of the conclusions of this work through periodic evaluations (studies) of the airport choice in Sao Paulo Metropolitan Area.

In order to satisfy the level of importance placed by air passengers on the access time, when the market is segmented by flight duration, this work advises the Air Transportation Authority to schedule short-haul flights at CGH, close to the center of origin of the air transport demand in Sao Paulo Metropolitan Area. In this way, long-haul flights should therefore be scheduled at GRU.

Meanwhile, to establish a perimeter rule in the air transport regional system in Sao Paulo Metropolitan Area, besides this passenger market analysis, it is also necessary to conduct a technical evaluation of the landside and airside capacities of the airports where the departures and connections would be scheduled (the latter may involve airports outside the region), as well as an economic evaluation from the standpoint of the yield management in case a perimeter rule is to be adopted.

For the airport managers, it is recommended to reserve areas at their terminals for business, social and cultural events, which should be advertised countrywide, attracting residents and visitors (non-residents). While the residents are potential passengers, the visitors enhance their experience with the airport where the event is held, factor revealed in this work to be best explaining the airport choice in Sao Paulo Metropolitan Area.

Since an expressive amount of passengers either declared having chosen their airport of departure not being aware of the supply of flights to their desired destination from a competing airport located in the same metropolitan area or even someone else (possibly somebody from their workplace or a travel agent) made their airport choice for them, therefore the events made available in a competing airport is a potential policy to broaden the participants minds in terms of considering this airport as an alternative of departure for their following flights.

This work showed that the choice of GRU is not base on the rationale of access time savings and increase in the supply of direct flight frequencies in the period of departure to the desired destination. The rationale of choosing GRU is based on the passenger experience with the airports in Sao Paulo Metropolitan. Thus, it is recommended that the airlines offer non-aeronautical advantages to make a passenger choose GRU as an airport of departure for the analyzed domestic destinations. For instance, these advantages may be related to the provision of either ground transportation to access GRU or a parking lot in the vicinity of GRU.

Since the experience was found to be the factor best explaining the airport choice in Sao Paulo Metropolitan Area, this work points out that the airlines can increase their aircraft occupancy by concentrating marketing efforts on those passengers who place more importance on their experience with the two airports: those who depart from CGH and those who fly shorter distances.

Because it was found that the choice of CGH is motivated by the proximity in terms of the variable of access time, while the choice of GRU is not, this work points out for the urban transportation planners that an improvement on the access such as connecting GRU to the subway system existing in the city of Sao Paulo, is likely to enhance the choice of CGH in case a station is included in the city of Guarulhos, rather than enhance the choice of GRU by the passengers living in the city of Sao Paulo (for the competing domestic flights).

8.Future work

Three alternatives are proposed to extend the research on airport choice in Sao Paulo Metropolitan Area:

a) Capturing possible changes in passenger choice behavior in Sao Paulo Metropolitan Area in different months of the year;

b) Applying the very same principles employed in this paper to other brazilian multiple airport regions;

c) Applying the very same principles employed in this paper to other air transport related choices in Sao Paulo Metropolitan Area, such as airline choice for domestic and international flights and the choice of interurban transportation mode, in which the air mode competes against the ground mode for several destinations.

9. References

- Ashford N. and M. Benchemam, (1987), Passengers' Choice of Airport: An Application of the multinomial Logit Model, *Transportation Research Record* 1147, pp.1-5.
- Augustinus, J. G. and S. A. Demakopoulos, (1978), Air Passenger Distribution Model for a Multiterminal Airport System, *Transportation Research Record* 673, pp.176-180.
- CET (Companhia de Engenharia de Trafego) (Traffic Engineering Company), (2000), Desempenho do sistema viário. (Traffic system performance).
- CET (Companhia de Engenharia de Trafego) (Traffic Engineering Company), (2001), Informações sobre tráfego e transportes. (Information on Traffic and Transports)
- Econometric Software, Inc., (2002), Reference Guide NLOGIT Version 3.0.
- Harvey, G. (1987), Airport Choice in a Multiple Airport Region, *Transportation Research A*, volume 21A (6), pp. 439-449.
- Hensher, D. and Q. Dalvi, (1978), Determinants of travel choice, Saxon House.
- Innes, J. D. and D. H. Doucet, (1990), Effects of Access Distance and Level of Service on Airport Choice, *Journal of Transportation Engineering*, volume 116 (4), pp. 507-516.
- Kanafani, A. (1983), *Transportation demand analysis*, McGraw-Hill.
- Koppelman, F. S. and C. Chu, (1985), Effect of sample size on disaggregate choice model estimation and prediction, *Transportation Research Record* 944, pp. 60-69.
- Moreno, M. B. and C. Muller (2002), Escolha de aeroporto em região de múltiplos aeroportos: o caso da Grande São Paulo (Airport choice in a multiple airport region: the case of Sao Paulo Metropolitan Area), MSc Dissertation, ITA – Instituto Tecnológico de Aeronautica (Technological Institute of Aeronautics), Brazil.
- Pels, E., P. Nijkamp and P. Ritveld, (1998), Airport choice in a multiple airport region: an empirical analysis for the San Francisco Bay Area, discussion paper 98-041/3, Tinbergen Institute, Holand.
- Skinner Jr, R. E. (1976), Airport Choice: An Empirical Study, *Transportation Engineering Journal of ASCE*, volume 102 (TE4), november, pp.871-882.
- Windle, R. and M. Dresner, (1995), Airport Choice in Multiple-Airport Regions, *Journal of Transportation Engineering*, july/august, pp. 332-337.