EFFECT OF SAMPLE STRATIFICATION STRATEGIES ON AIRPORT CHOICE EXPERIMENTS

Marcelo Baena Moreno
Energyarc Industrial Limitada

ABSTRACT
Using the conditional LOGIT model, this paper addresses the effect of sample stratification strategies on airport choice experiments. The explanatory power of twelve level-of-service variables is analyzed, considering three sampling strategies: random sample, sample stratified by total weekly flights to each destination and sample stratified by observed traffic. Irrespective of the stratification strategy, the results show that an increase in the flight frequency increases the attractiveness of an airport, as expected. Indirect flights (connections or stops) had the worst explanatory power on airport choice in comparison with the direct flights, also an intuitive result. The best explanatory power was shown by the models including direct flight frequencies when the sample was stratified by observed traffic. Therefore future samples should be planned in this way.

RESUMO
Usando modelo LOGIT condicional, este artigo endereça o efeito de estratégias de estratificação de amostras para experimentos de escolha de aeroporto. O poder explanatório de doze variáveis de nível de serviço foi analisado, considerando três estratégias de amostragem: amostra aleatória, amostra estratificada por vôos semanais totais para cada destino e amostra estratificada por tráfego observado. Independentemente da estratégia de estratificação, os resultados mostram que um aumento na freqüência de vôos aumenta a atratividade de um aeroporto, como esperado. Vôos indiretos (conexões e paradas) tiveram o pior poder explanatório sobre a escolha de aeroporto em comparação com vôos diretos, também um resultado intuitivo. O melhor poder explanatório foi mostrado pelos modelos que incluíram freqüências de vôos diretos quando a amostra foi estratificada por trafego observado. Portanto amostras futuras devem ser planejadas desta forma.

1. INTRODUCTION

According to the air transport long-term rising trend, there is a strong likelihood of a steady increase in the number of multiple-airport regions worldwide. This likelihood has prompted this research on the tradeoffs passengers face when flying out of multiple-airport regions.

"Airport choice research" constitutes an important field of transportation planning and is aimed at giving planning insight to professionals directly or indirectly involved with the air transportation activity, such as the air transportation authority, airport managers, airline managers and urban transportation planners.

The overwhelming majority of researchers on airport choice found that two kinds of variables satisfactorily explained 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.

Considering three sample stratification strategies, this paper aims at electing one to be adopted for a future airport choice study, measuring the explanatory power of level-of-service variables on the airport choice made by the passengers whose travel starts in Sao Paulo Metropolitan Area.

The region under 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 thanks to their volume of passengers and they compete with each other for the air passengers flying to numerous domestic destinations.

2. LITERATURE

In the past decades, airport choice has established itself as a relevant field of research in transportation planning. Papers on this theme include Skinner (1976), Augustinus and Demakopoulos (1978), Harvey (1987), Ashford and Benchemam (1987), Innes and Doucet (1990), Windle and Dresner (1995), Pels *et al.* (1998), Basar and Bhat (2002), Hess and Polak (2003), Moreno and Muller (2003), Moreno and Muller (2004). Moreno (2005) presents a detailed review of these papers.

Several multiple-airport regions worldwide have already been analyzed: Baltimore-Washington, New York-New Jersey, San Francisco Bay Area in the USA, Central England in the UK, New Brunswick in Canada and Sao

Paulo in Brazil. In these studies the modeling tools may have varied; however, a model of the LOGIT family was consistently the most frequently adopted one.

By and large these researchers have used variables related to the accessibility to the competing airports and the air carrier's level-of-service available at these air terminals. Moreover, Windle and Dresner (1995) have introduced a variable related to the experience a passenger has with the airports in the region.

Since the evolution of passenger behavior across time indicates the demand for seasonal researches on airport choice in the same multiple-airport region, Mota and Silva (2003) have remodeled the airport choice in Sao Paulo Metropolitan Area, the region analyzed by Moreno and Muller (2003) and Moreno and Muller (2004). Gathering data in the second semester of the year, Mota and Silva (2003) found some differences regarding the passenger profile: higher percentage of business travelers and higher percentage of low-cost airline flyers. Despite these differences, the best access variable (access time) and the best level-of-service variable (direct flight frequency in the period of departure) were found to be the same, and therefore not influenced by the time of the year when the survey was conducted.

Except in the case of Pels *et al.* (1998) who accounted for the disparity between sample and population fractions of passengers at each airport, it would seem that most papers cited lack a clear-cut sample stratification strategy. Moreover, in most cases it is the airport authority that collects data and makes it available to researchers. Once again exception is made for some papers such as Moreno and Muller (2003), Moreno and Muller (2004) and Mota and Silva (2003), whose data collections were self-sponsored.

3. MODEL

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 inequation:

$$U_{in} \geq U_{jn} \; for \; all \; j, 1 \leq j \leq z$$

Where U_{in}: utility that passenger n obtains by choosing airline i
U_{jn}: utility that passenger n obtains by choosing airline j
z: 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:

$$U_{in} = V_{in} + \varepsilon_{in} \; for \; all \; i, 1 \leq i \leq z$$

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

$$p_{in} = \frac{e^{V_{in}}}{\sum_{j=1}^{j} e^{V_{jn}}}$$

Where p_{in}: probability of passenger n choosing alternative i

j: number of alternatives (airlines)
V_{in}: 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 variables whose value remain constant across alternatives (for the same passenger) are also allowed.

The latter is the model that this paper employs, also known as 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.

4. SAMPLE

Three sample stratification strategies were considered:

- Strategy 1: Random sample of 1923 observations.
- Strategy 2: Sample was scaled to 2000 observations, a round number close to the random sample, and was stratified by existing weekly total (direct plus indirect) flight supply (variable-based) per destination in February 2002, the period when the random sample above was collected, finally yielding 2002 observations.
- Strategy 3: Once again the sample was scaled to 2000 observations, but now it was stratified by observed traffic (choice-based) in the annual report of year 2000, the last annual report published to the date when the random sample above was collected, finally yielding 1996 observations.

Tables 1, 2 and 3 show the market segmentation of these three samples, by chosen airport, flight range and popularity of the destination.

The data on Strategy 1 have been collected as part of a research work on airport choice (Moreno, 2002). Although Sao Paulo Metropolitan Area groups several cities, seven of them (Sao Paulo, Guarulhos, Santo Andre, Sao Bernardo do Campo, Sao Caetano do Sul, Diadema and Osasco) have been elected to portray the trip origin in this region, because of two reasons: first, they represent 79% of the electric power consumption in the region; second, the calculation of the access time from the other cities wasn't likely to lead to sound values.

Table 1: Sample segmentation of Strategy 1

Segmentation criteria	PAX	Market segments
Departure airport	897	Departing from GRU
	1026	Departing from CGH
Flight duration	807	In short-haul flights (up to 1 hour)
	1116	In long-haul flights
Destination popularity	341	Flying to the most popular destination (Rio de Janeiro)
	1582	Flying to other destinations

Table 2: Sample segmentation of Strategy 2

Segmentation criteria	PAX	Market segments
Departure airport	600	Departing from GRU
	1402	Departing from CGH
Flight duration	788	In short-haul flights (up to 1 hour)
	1214	In long-haul flights
Destination popularity	302	Flying to the most popular destination (Rio de Janeiro)
	1700	Flying to other destinations

Table 3: Sample segmentation of Strategy 3

Segmentation criteria	PAX	Market segments
Departure airport	729	Departing from GRU
	1267	Departing from CGH
Flight duration	1078	In short-haul flights (up to 1 hour)
	918	In long-haul flights
Destination popularity	597	Flying to the most popular destination (Rio de Janeiro)
	1399	Flying to other destinations

The alternative airports whose choice was analyzed were Congonhas Airport (CGH, located in the city of Sao Paulo) and Guarulhos Airport (GRU, located in the city of Guarulhos, in Sao Paulo Metropolitan area). The criteria for destination selection were that there must have been departures to these destinations from both airports and that the annual traffic volume must have surpassed 100,000 passengers. This volume avoids small sample bias (Windle and Dresner, 1995), represented by less popular destinations. These two requisites were evaluated through the DAC (Department of Civil Aviation) annual report of the year 2000. Table 4 presents the 21 airports studied in this paper, which correspond to 19 cities (two of them are multiple-airport destinations).

Table 4: Destinations studied in this paper

Destination airport	City where each airport is located
BPS	Porto Seguro
BSB	Brasília
CGR	Campo Grande
CNF	Belo Horizonte
CWB	Curitiba
FLN	Florianópolis
FOR	Fortaleza
GIG	Rio de Janeiro
GYN	Goiânia
IGU	Foz do Iguaçu
JOI	Joinville
LDB	Londrina
NVT	Navegantes
PLU	Belo Horizonte
POA	Porto Alegre
RAO	Ribeirão Preto
REC	Recife
SDU	Rio de Janeiro
SSA	Salvador
UDI	Uberlândia
VIX	Vitória

For Strategy 1, 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 only during these periods by CET (Traffic Engineering Company), enabling the calculation of access time to the airports.

For Strategy 1, 1923 passengers were interviewed. This sample size is satisfactory taking into account Koppelman and Chu (1985) who calculated the amount of observations required for relatively simple disaggregate choice models.

5. EXPLANATORY VARIABLES

5.1. Selecting a variable for sample stratification purposes

Three types of variables could theoretically be tested: those associated with the ground accessibility to the airports, those related to the airlines' level-of-service available at the competing airports and one associated with the passenger experience with the airports in the studied region.

The access-related variables are trip origin zone-related. Therefore access-based sample stratification could be carried out scaling the air transport demand to the electric power consumption existing in each zone. The problem is that the airport choice behavior of these passengers is unknown a priori.

Yet, airlines' level-of-service variables are able to cope with this problem, since both flight frequencies and observed traffic result from the combination between passenger preference and availability. If passengers didn't prefer a specific flight, its seat factor would go down and therefore no profit would come out. In this case, the traffic to this destination would decrease as well as the flight supply.

Airport experience is unknown a priori, so this variable is not ideal to be used as a sample stratification method to plan data collection.

Therefore level-of-service variables were employed in this paper. Level-of-service was portrayed by twelve flight frequency-related variables. 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.

For Strategy 1, 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 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 weekly frequency irrespective of day and period; 5)IDPF: Indirect (with connections or stops) frequency in the travel day and period; 6)IDF: Indirect frequency in the travel day; 7)IPF: Indirect frequency in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled; 8)IWF: Indirect weekly frequency irrespective of day and period; 9)TDPF: Total (direct plus indirect) frequency in the travel day and period; 10)TDF: Total frequency in the travel day; 11)TPF: Total frequency in the travel period (morning or afternoon peak) in all days of the week when the passenger traveled; 12)TWF: Total weekly frequency irrespective of day and period.

As an example, consider a passenger that departed to a given destination in the morning peak of February 19th, 2002. The variables of direct flight frequency are shown in Table 5, and the variables of indirect and total flight frequencies were built in an analogous way:

Table 5: Variables of direct flight frequency

	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

For Strategies 2 and 3, data from the morning of February 25th were used for DDPF and DDF. For Strategy 2, TWF was used for the stratification purpose for two reasons: Firstly, some destinations are not directly reached by CGH, therefore using a variable of direct flight frequency for a sample stratification purpose would result in some destinations being underrepresented. Besides, indirect flights are expected to be disregarded in front of direct flights. Therefore a total flight frequency should be used. Secondly, it is not known a priori which variable pertaining to the group of total ones would show the best explanatory power on airport choice. Therefore the most general one should be used, the total weekly frequency (TWF).

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 departure airport because of 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, flight frequency to the other airport in the multiple airport region of destination was disregarded when building the several variables of flight frequency variables. 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. This obviously, only applied to the Strategy 1.

5.2. Model structure using level-of-service variables

Using the conditional LOGIT model, the utility function of an alternative was designed as the effect of a level-of-service variable 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 n, $1 \leq n \leq 1923$ (Strategy 1), $1 \leq n \leq 2002$ (Strategy 2), $1 \leq n \leq 1996$ (Strategy 3), has been represented by two generic decision functions, being the first always for GRU and the second for CGH, whichever the departure airport, as follows:

$V_{GRU\,n} = \alpha_1.A_GRU + \alpha_2.FREQUENCY_{GRU\,n}$

$V_{CGH\,n} = \alpha_1.0 + \alpha_2.FREQUENCY_{CGH\,n}$

Where
V: deterministic part of the utility function of each airport for passenger n
α_k: parameter (coefficient) related to the variable k ($k=1$ for A_GRU and $k=2$ for FREQUENCY)
A_GRU: dummy alternative-specific constant (ASC) for GRU
$FREQUENCY$: variable associated with the airlines' level-of-service

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

$$p_{GRUn} = \frac{e^{V_{GRUn}}}{\sum_j e^{V_{jn}}}$$

$$p_{CGHn} = 1 - p_{GRU\,n} = \frac{e^{V_{CGH\,n}}}{\sum_j e^{V_{jn}}}$$

Where
p_{GRUn}: probability that passenger n chooses GRU
p_{CGHn}: probability that passenger n chooses CGH
V: deterministic part of the utility function of each airport for passenger n

6. RESULTS

Twelve one-variable models were built for each hypothesis, using one by one the twelve airline level-of-service variables. The comparison of these models reveals the variable with best explanatory power on the airport choice in Sao Paulo Metropolitan Area. Tables 6, 7 and 8 present the calibration results of these models.

Table 6: Calibration results for Strategy 1

Variable	A_GRU Coefficient	t-Student	p-value	Variable Coefficient	t-Student	p-value	Correct Prediction
DDPF	0.333	6.0	0	0.226	17.0	0	59.4%
DDF	0.332	6.1	0	0.035	15.9	0	58.9%
DPF	0.303	5.4	0	0.038	17.3	0	59.7%
DWF	0.305	5.6	0	0.006	16.2	0	59.2%
IDPF	0.199	2.7	0	0.094	6.0	0	51.2%
IDF	0.454	5.8	0	0.029	9.3	0	52.6%
IPF	0.161	2.3	0.01	0.014	5.6	0	51.1%
IWF	0.419	5.3	0	0.004	8.7	0	52.3%
TDPF	0.812	10.9	0	0.165	16.7	0	58.5%
TDF	0.834	11.6	0	0.028	17.1	0	59.3%
TPF	0.767	10.5	0	0.027	16.7	0	58.5%
TWF	0.823	11.6	0	0.005	17.2	0	59.6%

The similarities among the results for the three strategies are:

1) All variable coefficients are positive, what is expected, since the wider the variety of flights, the higher the airport choice probability.

2) All level-of-service variable coefficients had zero as p-value, what is desired.

Table 7: Calibration results for Strategy 2

Variable	A_GRU Coefficient	t-Student	p-value	Variable Coefficient	t-Student	p-value	Correct Prediction
DDPF	0.705	13.2	0	0.219	14.2	0	64.2%
DDF	0.547	10.2	0	0.047	13.1	0	63.7%
DPF	0.749	14.0	0	0.033	14.5	0	64.2%
DWF	0.587	11.0	0	0.007	13.4	0	63.8%
IDPF	0.581	9.3	0	0.097	7.5	0	59.9%
IDF	0.359	4.9	0	0.028	9.0	0	60.5%
IPF	0.620	10.4	0	0.015	7.3	0	59.8%
IWF	0.370	5.0	0	0.004	8.8	0	60.4%
TDPF	0.349	5.9	0	0.132	14.6	0	63.7%
TDF	0.061	0.9	0.36	0.031	14.4	0	64.1%
TPF	0.422	7.4	0	0.021	14.5	0	63.7%
TWF	0.084	1.2	0.19	0.005	14.5	0	64.3%

Table 8: Calibration results for Strategy 3

Variable	A_GRU Coefficient	t-Student	p-value	Variable Coefficient	t-Student	p-value	Correct Prediction
DDPF	-0.142	-2.3	0.02	0.354	20.2	0	72.3%
DDF	-0.404	-6.3	0	0.074	19.1	0	71.2%
DPF	-0.070	-1.1	0.25	0.055	20.5	0	72.4%
DWF	-0.350	-5.5	0	0.012	19.4	0	71.8%
IDPF	0.117	1.7	0.08	0.118	8.6	0	55.5%
IDF	-0.049	-0.6	0.51	0.030	9.6	0	56.0%
IPF	0.140	2.1	0.03	0.019	8.8	0	55.5%
IWF	-0.018	-0.2	0.80	0.005	9.1	0	55.7%
TDPF	-0.645	-9.0	0	0.195	20.9	0	68.2%
TDF	-0.935	-12.0	0	0.041	20.6	0	68.3%
TPF	-0.546	-7.9	0	0.032	20.3	0	68.3%
TWF	-0.918	-11.8	0	0.007	20.4	0	68.7%

3) The average probability of correct prediction fell between 50% and 80% for all the variables, as an indication of at least medium quality adjustments. It must be reminded that in an airport choice modeling, other variables such as access time and experience with the analyzed airports are to be included, what would increase the average probability of correct prediction.

4) From the point of view of a connection or a stop on the way to the destination, the indirect flight frequencies presented lower average probability of correct prediction in comparison with the direct and total ones. 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.

The differences among the results for the three strategies are:

1) P-values were zero or close to zero for models built according to strategy 1. Yet models including TDF and TWF presented undesired high p-values for the ASCs, when built according to strategy 2. Regarding strategy 3, there were many models with undesired high p-values, what occurred mostly with indirect flights.

2) The modulus of t-student is higher than 2 for all models built according to strategy 1, what means a high level of confidence. For strategies 2 and 3, generally when the p-value was undesired, the t-student was also undesired.

3) The best explanatory power among the three hypotheses was shown by the variable DPF in strategy 3, with 72.4% of the data being explained by this model. The fact that the ASC coefficient in this case presented a high p-value does not eliminate its merit definitely, since two other variables should be included in an airport choice modeling, such as access time and experience, and the interaction among variables may produce a lower p-value for the ASC.

4) Strategy 1 didn't show the best explanatory power on airport choice in any case. Strategy 2 showed the best explanatory power on airport choice for all indirect flight frequencies. Strategy 3 showed the best explanatory power on airport choice for all direct and total flight frequencies.

5) Out of 12 models of strategy 1, the one with DPF stood out as the best one. Since this is the random sample used in Moreno (2002), it is opportune to say that Mota and Silva (2003) also encountered the best explanatory power of DPF for Sao Paulo Metropolitan Area in the year following Moreno (2002). Among the 12 models of strategy 2, the one with TWF showed best explanatory power on airport choice, coinciding with the variable selected for sample stratification. For strategy 3, the best variable was similar to strategy 1 (DPF).

6) For strategies 1 and 3, 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). Since two hypotheses support these outcomes, there is a strong likelihood that they are right, despite strategy 2 not confirming them.

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. To 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.

7) For strategy 1, among the indirect flight 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. In the case of strategy 3, it was found IDF, IWF, IPF and IDPF. 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/she 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. Likewise, in the case of hypothesis 2 IDF stood out, being followed by IWF.

8) For strategy 1, among the total flight 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 and TDPF. . In the case of strategy 2, it was found TWF, TDF, TPF and TDPF. In the case of strategy 3, it was found TWF, TPF, TDF and TDPF. Also within groups the difference was not so big.

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.

Three sample stratification strategies were employed: random sample, sample stratified by total weekly flights (variable-based) to each destination and sample stratified by observed traffic (choice-based).

Several specifications for the decision function were tested. These specifications enhanced independent variables associated to the airlines' level-of-service in the airports.

Regarding the empirical results, since the average probability of correct prediction was higher for strategies 2 and 3 when compared with strategy 1, then in future airport choice experiments sample stratification should be attempted. Since we know a priori that direct flight frequencies are far more appealing than indirect or total ones, then the sample stratification strategy should be that of strategy 3, as opposed to strategy 2, in which some total flight frequencies performed better than the corresponding direct ones.

From the theoretical point of view the above conclusions are also supported. Stratifying the sample by total weekly flights constitutes an improvement over no stratification at all, since it accounts for the flight supply. However, aircrafts do not flying totally full of passengers, and thus stratifying the sample by observed traffic is the best way to account for the load factor, since it represents the population proportions of passengers flying to each destination. As a result, it enables the larger percentage of passengers flying to the most popular destinations be correctly represented as opposed to less popular destinations, despite of flight frequency.

Finally, the specification that produced the models most adjusted to the data (evaluated in terms of the highest average probability of correct prediction) enhanced direct flight frequencies, when the sample was stratified by observed traffic.

Therefore, 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 in future airport choice experiments, especially those self-sponsored, the sample should be stratified by observed traffic.

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.

Basar, G. and C. Bhat (2002) *A Parameterized Consideration Set Model for Airport Choice: An Application to the San Francisco Bay Area*, Technical Paper, Department of Civil Engineering, University of Texas at Austin.

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.

Hess, S. and J. W. Polak (2003) *Mixed Logit modeling of airport choice in multi-airport regions*, Working Paper, Centre for Transport Studies, Imperial College London.

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. (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 Aeronáutica (Technological Institute of Aeronautics), Brazil.

Moreno, M. B. and C. Muller (2003) *Airport choice in Sao Paulo Metropolitan Area: an application of the conditional LOGIT model*, 7th Air Transportation Research Society (ATRS) World Conference, Toulouse, France.

Moreno, M. B. and C. Muller (2004) *Airport choice in Sao Paulo Metropolitan Area: an analysis based on passenger market segmentation*, Air Transportation Research Society (ATRS) World Conference, Istanbul, Turkey.

Moreno, M. B. (2005) *Influence of passenger experience on airport choice in Sao Paulo Metropolitan Area*, 4th Air Transport Symposium (SITRAER), Sao Jose dos Campos, Brazil

Mota, D. C. and M. V. C. M. Silva (2003) *Remodelamento do processo de escolha de aeroportos na Grande São Paulo* (Remodeling the airport choice process in the Sao Paulo Metropolitan Area), Civil Engineering Final Graduation Research, ITA – Instituto Tecnológico de Aeronáutica (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.

---

caixamarcelo@ig.com.br