INSTITUTO TECNOLÓGICO DE AERONÁUTICA

Paulo Thiago Araujo Moraes

Application of a nested logit model for identification of potential demand in regional air transportation markets

Trabalho de Graduação

2011

Civil

CDU: 656.7.02

Paulo Thiago Araujo Moraes

APPLICATION OF A NESTED LOGIT MODEL FOR IDENTIFICATION OF POTENTIAL DEMAND IN REGIONAL AIR TRANSPORTATION MARKETS

Orientador
Prof. Dr. Alessandro Vinícius Marques de Oliveira (ITA)

Engenharia Civil-Aeronáutica

SÃO JOSÉ DOS CAMPOS
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA AEROESPACIAL
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2011

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

Moraes, Paulo Thiago Araujo
Application of a nested logit model for identification of potential demand in regional air transportation markets/ Paulo Thiago Araujo Moraes
São José dos Campos, 2011.
49f.

Trabalho de Graduação – Divisão de Engenharia Civil – Instituto Tecnológico de Aeronáutica, 2011. Orientador: Prof. Dr. Alessandro Vinícius Marques de Oliveira

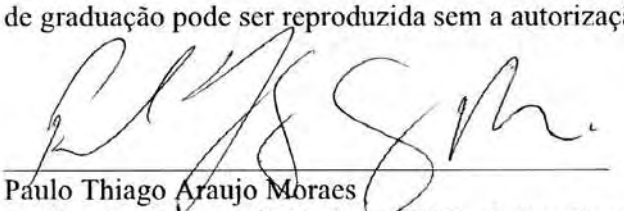
1. Air Transportation. 2. Discrete choice model 3. Nested Logit. I. Departamento de Ciência e Tecnologia Aeroespacial. Instituto Tecnológico de Aeronáutica. Divisão de Engenharia Civil. II. Application of a nested logit model for identification of potential demand in regional air transportation markets

REFERÊNCIA BIBLIOGRÁFICA

MORAES, Paulo Thiago Araujo. Application of a nested logit model for identification of potential demand in regional air transportation markets. 2011. 49f.
Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Paulo Thiago Araujo Moraes
TÍTULO DO TRABALHO: Application of a nested logit model for identification of potential demand in regional air transportation markets
TIPO DO TRABALHO/ANO: Graduação / 2011



Paulo Thiago Araujo Moraes
Av. Santos Dumont 6944, Apto 603/A, Bairro Papicu, Fortaleza - CE

APPLICATION OF A NESTED LOGIT MODEL FOR IDENTIFICATION OF POTENTIAL DEMAND IN REGIONAL AIR TRANSPORTATION MARKETS

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação

Paulo Thiago Araujo Moraes
Autor

Prof. Dr. Alessandro Vinícius Marques de Oliveira (ITA)
Orientador

Prof. Dr. Eliseu Lucena Neto
Coordenador do Curso de Engenharia Civil-Aeronáutica

São José dos Campos, 21 de Novembro de 2011

Dedico este trabalho a Deus,
por ter permitido que eu chegasse até aqui,
a meus pais, Gerardo e Edilene, e a minha irmã, Emília

Agradecimentos

Agradeço a Deus, por ter me dado forças nesses cinco anos para suportar todas as adversidades que surgiram.

Agradeço ao meu Pai, Gerardo, pela dedicação e paciência que sempre teve comigo e com minha irmã.

Agradeço à minha Mãe, Edilene, pelas inúmeras orações e por ter me ensinado a confiar em mim mesmo e em Deus.

Agradeço à minha irmã, Emília, pelo apoio fundamental nesses últimos cinco anos e por sempre ter sido uma pessoa amável e dócil.

Agradeço ao Professor Alessandro pela orientação nesse trabalho.

Agradeço também a todos os bons e péssimos professores que tive durante toda a minha vida, por terem sido fundamentais em diversas escolhas profissionais e pessoais.

Aos amigos do apartamento 223 deixo o meu muito obrigado pelos cinco fantásticos anos que tivemos juntos. Agradeço ao Renato (Ratão), por toda a sua paciência; ao Armando (Asqueroso), por ter suportado todas as brincadeiras sem graça durante tanto tempo; ao Diniz pelas boas opiniões sobre livros; ao Gilberto (Madruga), dupla de LAB durante todo o ITA, pela ajuda nas inúmeras provas; ao Fábio (17), com quem tive o prazer de dividir o quarto nesses 5 anos, pelas incontáveis boas risadas que demos juntos; ao Pedro (Fiat Lux), por ter se integrado ao nosso apartamento trazendo boas histórias; ao Samuel, pelos ótimos anos de convivência no FUND e na INFRA; ao Rômulo (Brasil), pela presença constante em nosso apartamento.

Aos amigos da INFRA, posso dizer que realmente vivemos bons momentos juntos entre aulas, visitas e viagens. Devo um agradecimento especial a Mayara, pelo seu caderno e pelas ajudas em vésperas de provas, ao Bruno (Pinky) e ao Renato (Tubies) por tantas caronas, almoços, vésperas de provas e alguns viradões juntos.

Aos outros inúmeros amigos de H8, agradeço por terem feito parte da minha passagem por aqui. Vocês certamente fizeram o ITA valer à pena.

Agradeço também a todas as pessoas que de alguma forma cruzaram meu caminho até esse momento e que contribuíram para que eu chegasse até aqui.

“Porque aos seus anjos dará ordens ao teu respeito,
Para que te guardem em todos os teus caminhos”

Salmo 91:11

“You gain strength, courage and confidence
By every experience in which you stop to look the fear in the face.
You are able to say to yourself
"I lived through this horror,
I can take the next thing that comes along"”

Eleanor Roosevelt

Resumo

Apoiado em uma situação econômica favorável e nas suas grandes dimensões, o Brasil necessita incluir nos estudos que norteiam as suas políticas públicas, questões de identificação de potencial demanda para o transporte aéreo. A cobertura do transporte aéreo ao longo do território nacional caiu ao longo da última década, sendo que um conjunto expressivo de cidades deixou de ser servido pela aviação regular. O presente trabalho mostra um modelo baseado em modelos de escolha discreta visando o apontamento de cidades com potencial de serem incluídas nas malhas das companhias aéreas regionais. Uma comparação entre os modelos tradicionais Probit e Logit sugere que são equivalentes. Utilizou-se também um modelo Logit Aninhado com características não observáveis visando elencar microregiões brasileiras com potencial de operação sustentável do ponto de vista econômico. Os resultados foram apresentados considerando o cenário sócio-econômico de 2008.

Abstract

Supported by a favored economical situation and its large dimensions, Brazil needs to include in studies that guides the public policies, issues of identification of potential demand for air transportation. The coverage of air transport throughout the country has dropped over the last decade, with a significant set of cities no longer served by regular aviation. The present work shows a model based on discrete choice models aimed at pointing to the potential of cities to be included in the meshes of regional airlines. A comparison between the traditional Probit and Logit models suggests that they are equivalent. A model Nested Logit with unobserved characteristics was used to rank the Brazilian microregions with potential sustainable economical operations. The results were presented considering the socio-economic scenario of 2008.

.

LIST OF FIGURES

LIST OF TABLES

Contents

1. Introduction

Given the favored situation of the Brazilian economy and its large territorial dimensions, air transportation is a major factor for a sustainable development of Brazil. Aviation in the country has experienced a consistent growth over the past years. Given by this scenario the design of a model for identification of the potential demand for regional air transportation is necessary for both public policy and infrastructure investment planning.

The model is based on discrete-choice theory. The basic assumptions is that airlines – The decision maker – choose to operate different domestic locations – the alternatives – according to attributes like economic activity, population, distance to the next airport hub, etc. I employ the two most traditional models of this kind Probit and Logit, perform a systematic comparison between them in terms of regional air transportation markets identification. In terms of the definition of locations, we consider expanded metropolitan areas ("microregions") as proxy of for the potential airport catchment area of cities.

The contribution of this study relies on the development of a Nested Logit model with unobserved characteristics suggested by Berry (1994). With respect to previous literature, the present work is certainly the first application of this sort of model in the study of potential demand in regional airline industry. The proposed discrete-choice framework has the following characteristics. First, it involves, in the top level, taking into account two basic alternatives for the airlines the "Inside Good" – ie, effectively operating a given aircraft-hour in the market, and the "Outside Good" – ie, letting that aircraft-hour grounded. Second, in the intermediate level of the decision-making tree of airlines, we model the choice between the 5 regions of Brazil – North, Notheast, Southeast, South and Midwest "macroregions" - , as we believe that locations within the same geographic region are similar to locations pertaining to different regions. And finally, the bottom level of the Nested Logit tree consists of the available locations to airlines. In order to pinpoint potential markets, I consider all possible locations, not only those with airports ready for schedule air transportation operations. In a version of empirical model, I consider only non-state capitals to check the robustness of results.

Finally, the proposed model was developed along with estimatios of Outside Good and Inside Good. Besides, the econometric modeling, including the choice of variables, database used, etc, are described and employed in the empirical framework.

The estimated equation of the Nested Logit model, based on the share that the flights in a microregion represent over the total flights in Brazil, is presented and then applied to empirically study the potential demand of Brazilian microregions.

2. Regional Air Transportation

2.1. Introduction

Brazil is a continental country with an area of approximately eight millions of squared kilometers and with more than five thousand cities. In 2010, it has experienced an economic growth of 7,5%, following the IBGE (Brazilian Institute of Geography and Statistics) estimation, what is impressive due to the proximity of the 2008 crisis.

So, due to Brazil's area and favored economical situation the air transportation should have a major role in this scenario. From the public policy standpoint, it is crucial to promote the development of the countryside of the country, especially in regions with poor economic and social indicators, aiming at reducing concentration of population and economic activity in few major cities. Thus, a strong regional air transportation system should support the development of the Brazilian countryside.

Besides, we nowadays have a situation of underinvestment and shortage of capacity of the major Brazilian airports which created barriers to the sustainable development of aviation in the main Brazilian cities.

Based on this, the regional segment of air transportation in Brazil has good opportunities for rapid growth, with regional airports demanding public and private investments in order to accommodate this dynamics.

2.2. Brazilian Aviation Situation

The air transportation in Brazil has increased since the beginning of the 2000's. Some changes in the Brazilian regulatory system in the end of the 1990's and the start of airlines with a Low Cost, Low fare business (e.g. GOL Airlines) caused several changes in the dynamics of the market.

The next Figures present the evolution of the air transportation in Brazil between 2006 and 2008. It is possible to notice that the Brazilian fleet, frequencies of flights, available seats and hours of flight presented a considerable growth.

Figure 1 – Commercial Airplanes Fleet

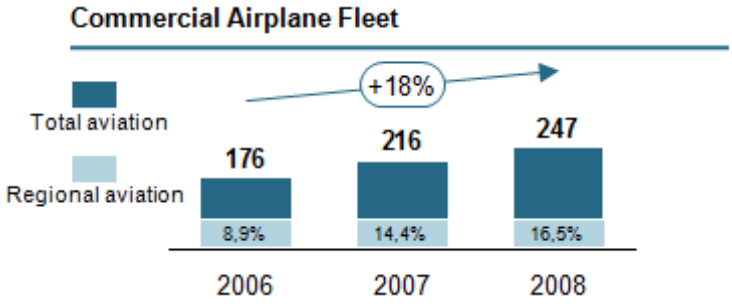


Figure 2 – Frequencies of Flights

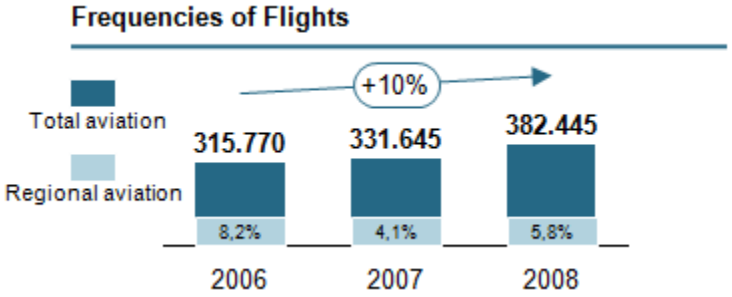


Figure 3 – Available Seats

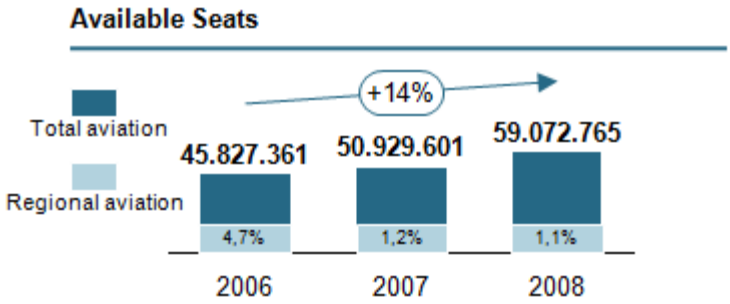
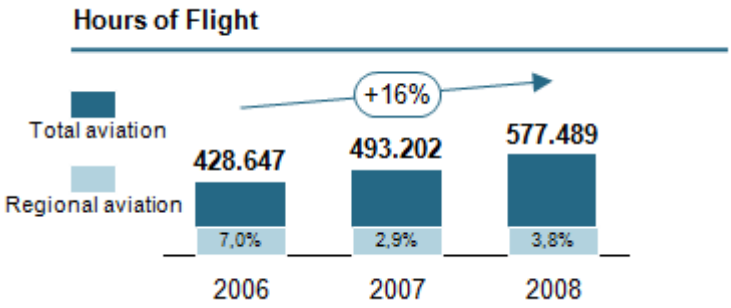


Figure 4 – Hours of Flight



Regarding regional air transportation in Brazil, the Figure 5 shows the total number of microregions per region followed by number of microregions with domestic commercial flights in 2008.

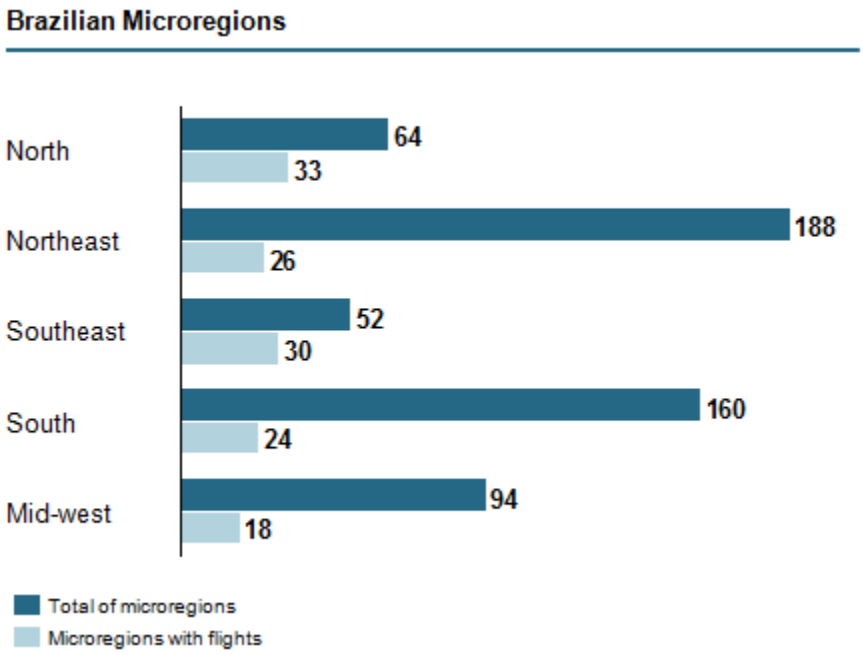


Figure 5 – Brazilian Microregions

The graphic above shows that the North region has approximately 50% of share of microregions with flights, probably due to the lack of highways and infrastructure for automobiles transportation. The Northeast region has roughly 12% of share only and Southeast microregions roughly 60% of share. Those differences occur probably because of the different economic development of regions and the different transportation infrastructure prioritized by the states of each region.

3. Discrete Choice Models

3.1. Introduction

Discrete choice models are conceived to represent the decision maker choosing one option among a set of possible alternatives. Those models are originally based on binary decisions with the following assumptions.

(1) Alternatives need to be mutually exclusive
(2) Alternatives must be exhaustive
(3) The number of alternatives must be finite
(4) Only one alternative is chosen, the one with highest utility

Those types of models are also derived from the random utility model (RUM) framework in which a utility is associated to each choice. This utility depends on the alternative and also on the decision maker and a decision never gives the same utility for different individuals.

Equation 1 shows the utility obtained by the individual *n* when the alternative *j* is chosen. It is composed by term that represents the observable factors that influence the utility, V_{nj} , and the term that considers the non-observable factors, ε_{nj} , also known as random term.

$$U_{nj} = V_{nj} + \varepsilon_{nj}, \forall n, j$$

Equation 1

To sum up, it can be inferred that the decision maker will choose the alternative that gives the highest utility and the value of utility obtained depends on intrinsic characteristics of each individual and the different characteristics of the alternatives. So, two individuals can make the same choice, but the utility is different for each one.

This work will present a short comparison between two important discrete choice models named Probit and Logit. The following topics will evaluate its behavior in a choice model for air transportation.

3.2. Probit and Logit

3.2.1. Probit

The Probit is one of the commonly used models to obtain binomial responses. As a discrete choice model, it is ideal to responses like 0 or 1, which may represent qualitative responses like yes or no, true or false etc. It is based on the utility obtained by the decision maker, as was discussed in the last topic.

The main difference of this model to others is the random part ε_{nj}, which is based in normal distribution that makes the Probit be the inverse cumulative distribution function of the normal distribution.

The following Equation 2 shows the function that defines a probability to a Probit model. One of the main characteristics is the fact that it does not have a "closed form", it has to be defined by the integral presented below.

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-x^2/2} dx$$

Equation 2

This equation results in a sigmoid curve, which is also known as "S-Curve" due to its shape and, often, is used to describe growths. By its shape, presented on the Figure 6 is possible to notice that in the beginning the growth is slow and rapidly become fast until arrive in the maturation point.

3.2.2. Logit

In the same way as Probit, Logit is also a discrete choice model to obtain binomial responses. It is based in the utility obtained by the decision maker when the choice is made and it is derived from a logistic distribution that has a form similar to the Equation 4.

The utility model is the same presented in the Equation 1 and the main difference to the Probit is the random part, ε_{nj}, which is based in a Gumbel distribution. Besides, the difference of two Gumbels distributions is distributed as a logistic, thus is possible to have the cumulative distribution function with following steps, where *n* represents a decision maker and *j* and *k* are different options.

$$P_{nj} = \Pr(U_{nj} \geq U_{nk}) = \int_{-\infty}^{\varepsilon_n} f(\varepsilon_n) = \int_{-\infty}^{\varepsilon_n} \frac{\mu e^{-\mu\varepsilon_n}}{(1+e^{-\mu\varepsilon_n})^2} d\varepsilon_n$$

Equation 3

$$P_{nj} = \frac{1}{1+e^{-\mu\varepsilon_n}}$$

Equation 4

Thus,

$$P_{nj} = \frac{1}{1+e^{-(V_{nj}-V_{nk})}} \frac{e^{V_{nj}}}{e^{V_{nk}}} = \frac{e^{V_{nj}}}{e^{V_{nj}}+e^{V_{nk}}}$$

Equation 5

So generalizing it to any options, the equation of the probability is given by

$$P_{nj} = \frac{e^{V_{nj}}}{\sum_i e^{V_{ni}}}$$

Equation 6

The Equation 6 represents de cumulative distribution function to the Logit model. This equation is a “closed form” equation, so it is most familiar to be used.

The curve obtained by the use of the Logit is also a sigmoid, but is not the same as the Probit's. Figure 6 shows a comparison with the curves of Probit and Logit and is possible to notice that the main difference of the curves is the begging and in the end of the evolution of the probability. So, usually is said that the Logit has “fattered” tails because of its shape quite different of Probit's being “slower” in the beginning and “faster” in the end.

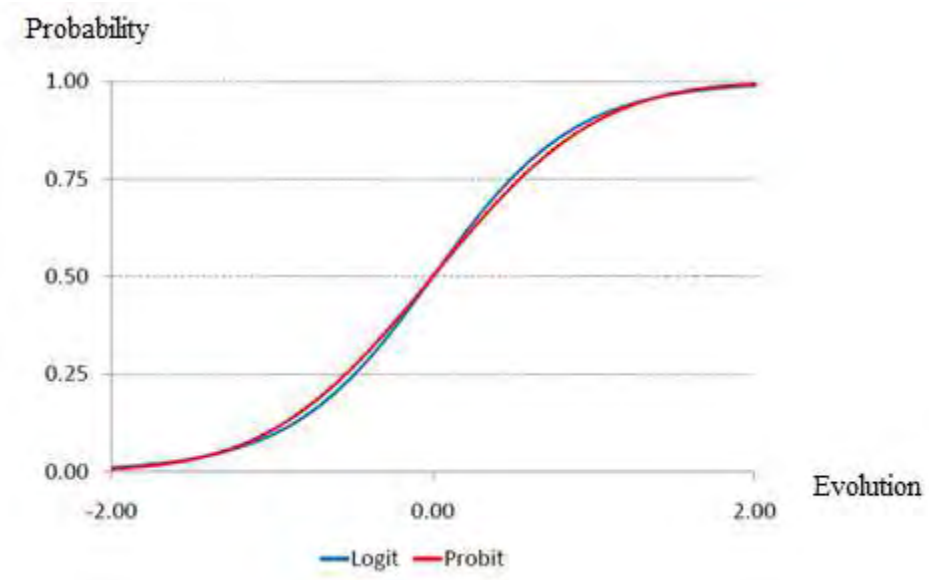


Figure 6 – Logit and Probit curves

3.2.3. Comparison

Logit and Probit are similar distributions due to the characteristics presented in the last topics. So, to introduce an example of comparison, a potential demand model will be presented in order to compare and try to evaluate them.

To create this, the fact of Brazil is divided in regions and each region is divided in microregions following the definitions and division made by IBGE (Brazilian Institute of Geography and Statistics) was considered. The main idea is to evaluate the possibility to exist a demand for regional air transportation in each of those microregions.

The model that will be shown is a Probit model that was developed by Oliveira and Salgado (2008). The dependent variable is called Pr[Probit] and is a 0 or 1 variable. 0 means that there is no potential demand and 1 means that certainly there is potential demand for regional air transportation. The model was applied to all the microregions in Brazil.

The Equation 7 shows the model with the independent variables showed below:

- Pr[Probit]: Represents the probability to a location have regular flights using the Probit model
- $\Phi[*]$: Function representative of the distribution of probabilities (Probit model)
- GDP_k: Gross domestic product of the microregion k, in BRL. In the model : sr_gdp
- $GDP\_boundary_k$: Gross domestic product of other microregions in the same mesoregion which k belongs, in BRL. In the model: xmr_gdp
- $Area_k$: Microregion area, in km². In the model: sr_area
- $Attractions_k$: Number of touristic attractions in the microregion k. In the model: sr_attr
- $Capital\_dist_k$: Average distance, in km², to the state capital of the cities that belong to the microregion k. In the model: km_med
- $Airport\_dist_k$: Distance to the closest airport in km. In the model: km_med
- $\Omega_k : reg\_co,\ reg\_ne,\ reg\_n,\ reg\_s$: Group of Dummies for regions Midwest, Northeast, North and South that represents a set of unobservable factors that may indicate potential demand. In the model: reg_co, reg_ne, reg_n, reg_s.
- ε_k: Regression random term

$$\Pr[\text{Probit}] = \begin{bmatrix} GDP_k, GDP\_boundary_k, \text{Area}_k, Attractions_k \\ Capital\_dist_k, Airport\_dist_k, \Omega_k, \varepsilon_k \end{bmatrix}$$

Equation 7

The presented model was a base to develop a series of comparisons between Logit and Probit, once the models are similar. The comparison was made using a group of variables to create two models similar to the presented one being a Probit and Logit.

However, in order to compare the two models with the same group with variables it was necessary to establish criteria. The following three criteria were used:

- R^2
- P-value
- Elasticity

The R^2 is a parameter to evaluate how good a curve fits to a set of data. Usually the R^2 has a value between 0 and 1, where 0 means the curve has no fit being the worst case and 1 means the curve has a perfect fit, the best case.

There are different ways to evaluate the R^2 of a curve and in this work the R^2 chosen was the McFadden's. A characteristic of the McFadden's R^2 is the fact it never reaches 1 and a value between 0,2 and 0,4 means a good fit of the curve to the data points. The Equation 8 shows how the McFadden's R^2 is calculated.

$$R^2 = 1 - \left(\frac{\ln L_A}{\ln L_0} \right)$$

Equation 8

Where,

- $\ln L_A$: Log – Likelihood of the alternative model. It is the likelihood considering a model with the coefficients estimated.
- $\ln L_0$: Log – Likelihood of the zero model. It is the model without the predictors.

The main purpose of this criteria is to evaluate whether the model (Probit or Logit), with the group of chosen variables, has a good fit or not. The R^2 was calculated for each case as will be shown.

The P-value is a way to see if a variable is statistically significant or not. In other words, that is a manner to evaluate if a variable will be rejected or not in the model and a rejected variable will not be part of the final model's equation.

This is good to see if the Probit and the Logit have the same behavior in rejecting or not a variable. The test used was based in three levels of significance 1%, 5% and 10%, in other words every time that the P-value is higher than one on those levels is said that the variable is rejected

The elasticity of a variable is a way to know how a change in this independent variable affects the dependent variable. The value of the elasticity means how a change of 1% in the dependent variable changes in the dependent.

For example, if y and x are variables in a linear regression where y is the dependent and x is the independent, a elasticity of 0,2 for y means that when the value of x is changed in 1% the value of y will have an increase of 0,2. Equation 9 shows the definition of elasticity.

$$\varepsilon_{y,x} = \frac{dy}{dx}$$

Equation 9

The purpose to calculate the elasticity for all the variables in the two tested models (Probit and Logit) is to understand the behavior of each term in the regression and try to see the level of importance of each variable in the models, once a variable with high elasticity it is supposed to be important and this fact needs to be understood.

The comparison between the models Probit and Logit used different groups of variables which are listed in each one of the five cases. Tables 1 to 3 shows the list of characteristics of each model that were observed each column has the following meaning

- *Variable*: Name of the variable used
- *Description*: Meaning of the variable
- *Coefficient*: Value of the coefficient in the regression
- *Value-p test result*: P-value result for each variable, where following code indicates the level of significant for which it is not rejected. The Coefficients are repeated.
    - *: P-value < 0,01 (1% of significance);
    - **: P-value < 0,05 (5% of significance);
    - ***: P-value < 0,1 (10% of significance);
- *Elasticity:* Elasticity, extracted at the mean of each variable
- *R^2:* McFadden's R^2 value

Case 1:

Table 1 - Probit and Logit results for Case 1

Model	Variable	Coefficient	P-value test Result	Elasticity	R^2
Probit	sr_gdp	2.10E-07	2.104E-07***	1.0914570	0.446
	xmr_gdp	-1.18E-08	-1.178E-08***	-0.2561047	
	sr_area	0.0000186	0.0000186***	0.3818132	
	sr_attr	0.0124051	0.01240514***	0.1224186	
	sr_kmnohb	0.0021457	0.00214574***	0.5052784	
	km_med	0.0082188	0.00821876***	1.2537670	
	reg_co	-0.3898189	-0.38981888	-0.0489688	
	reg_ne	-0.4257838	-0.42578384	-0.1933751	
	reg_n	0.0290084	0.0290084	0.0044849	
	reg_s	0.4817296	0.48172963**	0.1093918	
	_cons	-2.978325	-2.9783247***	-	
Logit	sr_gdp	5.41E-07	5.414E-07***	1.5679520	0.465
	xmr_gdp	-2.29E-08	-2.287E-08***	-0.2775576	
	sr_area	0.0000295	0.00002948***	0.3378780	
	sr_attr	0.0184552	0.01845519**	0.1016809	
	sr_kmnohb	0.0050148	0.00501478***	0.6592953	
	km_med	0.0145903	0.01459027***	1.2426510	
	reg_co	-0.5476987	-0.54769866	-0.0384126	
	reg_ne	-0.5963605	-0.59636051	-0.1512153	
	reg_n	0.3295693	0.32956933	0.0284483	
	reg_s	0.8806221	0.88062205**	0.1116468	
	_cons	-5.872517	-5.8725173***	-	

The Case 1 shows the same group of variables used by Oliveira and Salgado (2008) in the original Probit Model. The variables theoretically rejected (reg_co, reg_ne, reg_n) by the P-value test were the same in the Probit model and in the Logit Model.

Besides, Probit model showed a higher elasticity for the variable related to the GDP of the region. This is interesting because GDP is a very relevant variable related with economic potential of a microregion. Also, the R^2 for the two models was very good. In both cases the value was higher than 0,4 what shows a good fit of the curve to the regression obtained.

So, this first comparison is good to notice that none of the models compared had a great advantage in front of the other when comparing the quality of the regression obtained.

Case 2:

Table 2 - Probit and Logit results for Case 2

Model	Variable	Coefficient	P-value test result	Elasticity	R^2
Probit	sr_gdp	5.04E-08	5.043E-05	0.2745957	0.512
	xmr_gdp	-1.14E-08	-1.144E-08***	-0.2611135	
	sr_area	0.000016	0.00001615***	0.3479416	
	sr_attr	0.008518	0.00851801*	0.0882394	
	sr_kmnohb	0.003774	0.0037736***	0.9327988	
	km_med	0.008240	0.00823998***	1.3195190	
	reg_co	-0.417433	-0.4174331	-0.0550456	
	reg_ne	-1.065969	-1.0659686	-0.5082001	
	reg_n	-0.034202	-0.03420145	-0.0055508	
	reg_s	0.494798	0.49479799**	0.1179474	
	sr_pop	3.90E-06	3.897E-06***	1.8201230	
	_cons	-3.738218	-3.7382176***	-	
Logit	sr_gdp	1.40E-07	1.40E-04	0.4353386	0.509
	xmr_gdp	-2.13E-08	-2.134E-08***	-0.2787975	
	sr_area	0.000027	0.00002738***	0.3377897	
	sr_attr	0.015168	0.01516823*	0.0899520	
	sr_kmnohb	0.006764	0.00676383***	0.9571401	
	km_med	0.014357	0.01435674***	1.3161230	
	reg_co	-0.555320	-0.55531966	-0.0419209	
	reg_ne	-1.773307	-1.7733066	-0.4839779	
	reg_n	0.033565	0.03356513	0.0031185	
	reg_s	0.922941	0.922941**	0.1259464	
	sr_pop	6.51E-06	6.507E-06***	1.7397780	
	_cons	-6.635806	-6.635806***	-	

The Case 2 shows the same group of variables of Case 1 plus the population of each microregion. The reason for the addition of the population (represented by the population sr_pop) is to have two variables usually with high correlation: GDP and Population.

The variables rejected were also the same in both models, the elasticity's were also close in the Probit and in the Logit with some elasticity's higher in the Probit case and some higher in the Logit case. Besides, both R^2 were very close and in a range considered to have a good fit of the regression with data. So, this case also could not show any strong evidence that one of the models is better that the other.

Case 3:

Table 3 - Probit and Logit results for Case 3

Model	Variable	Coefficient	P-value test result	Elasticity	R^2
Probit	sr_gdp	1.90E-07	1.902E-07***	0.722549	0.391
	xmr_gdp	-1.01E-08	-1.008E-08***	-0.276363	
	sr_area	0.000019	0.00001932***	0.491191	
	sr_attr	0.010576	0.01057569**	0.105921	
	sr_kmnohb	0.002820	0.00281986***	0.857860	
	km_med	0.007292	0.00729198***	1.331896	
	reg_co	-0.552940	-0.55293987	-0.084589	
	reg_ne	-0.611457	-0.61145679	-0.341711	
	reg_n	-0.134925	-0.1349251	-0.024432	
	reg_s	0.434293	0.4342925*	0.123385	
	_cons	-2.917551	-2.9175506***	-	
Logit	sr_gdp	4.73E-07	4.733E-07***	0.956293	0.403
	xmr_gdp	-1.96E-08	-1.956E-08***	-0.285250	
	sr_area	0.000031	0.00003084***	0.416868	
	sr_attr	0.016544	0.01654391**	0.088122	
	sr_kmnohb	0.005744	0.0057443***	0.929388	
	km_med	0.013257	0.01325702***	1.287779	
	reg_co	-0.780772	-0.78077173	-0.063523	
	reg_ne	-0.914858	-0.91485761	-0.271905	
	reg_n	0.023711	0.02371093	0.002283	
	reg_s	0.816884	0.81688352*	0.123428	
	_cons	-5.637751	-5.6377506***	-	

The Case 3 shows the same group of variables used in the Case 1, but the database was modified to do not include the microregions with state capitals. That is interesting to understand the behavior of the regression when the main microregions were excluded.

The variables rejected were the same in the two models and also the same that in Case 1 regression. The elasticity of GDP, for example, was smaller in the Case 3 than in Case 1 what is reasonable once the GDP had its importance lowered because the higher GDP's were excluded. Besides that, the other elasticity's were also coherent with what was presented in Case 1.

The R^2 had a good value, close to 0,4 in both case, what shows the quality of regression's fit to the data.

The comparison done in the three cases presented results which were not conclusive to define whether Probit is better than Logit or the opposite, so the main conclusion that can be made looking to the data is that Probit and Logit are equivalent model with intrinsic characteristics.

It is necessary to understand that choose of Probit or Logit to develop a model is not possible to predict which one will generate the best results.

The following Figures shows the summary of the results presented on the last tables for the three cases analyzed.

Figure 7 – Results for Case 1

Case 1

	R^2	P-Value	Elasticities
Probit	0,446	Dummines CO, NE and N rejected	OK
Logit	0,465	Dummines CO, NE and N rejected	OK

Figure 8 – Results for Case 2

Case 2

	R^2	P-Value	Elasticities
Probit	0,512	Dummines CO, NE ,N and GDP rejected	PIB < POP
Logit	0,509	Dummines CO, NE ,N and GDP rejected	PIB < POP

Figure 9 – Results for Case 3

Case 3

	R^2	P-Value	Elasticities
Probit	0,391	Dummines CO, NE and N rejected	PIB >Others
Logit	0,403	Dummines CO, NE and N rejected	PIB >Others

4. Nested Logit

4.1. Introduction

The Nested Logit model is strongly based on the Logit, once they are described by almost the same assumptions. It can be explained as a Logit applied to different groups of data with some caveats and with some conditions and properties that make it different from other models.

As a discrete choice model, The Nested Logit is ideal to make choices. But the main characteristic of the model is to cluster the data in nests, which are groups of data with characteristics in common. Once the nests are created, the results for data in the same nest tends to be closer than the data in different nest, in other words the similar data generates more similar results. So, Nested Logit has the structure of nests and it needs to follow the next two assumptions.

- The ratio of probabilities of two alternatives in the same nest is independent of the existence of other alternatives in other nests.
- For alternatives in different nests the ratio of the probabilities can depend on other alternatives.

The following equation is the same presented on Equation 6 with the difference that now the options can be in more than one group, where the different groups are the nests. It is based on the utility model shown previously, where it shows the probability to choose the alternative *i* in the nest B_k, by the decision maker *n*. Being V_{ni} the observable part of the utility model described previously.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in B_k} e^{V_{nj}}}$$

Equation 10

An example of problem that can be modeled as a typical Nested Logit problem is the following. A person needs to go everyday to work but she has to choose the transportation modal. She can go by walking or by taking a ride with a friend and she can also go by own car, by bus or by train. Figure 10 shows a decision tree based on the five possibilities.

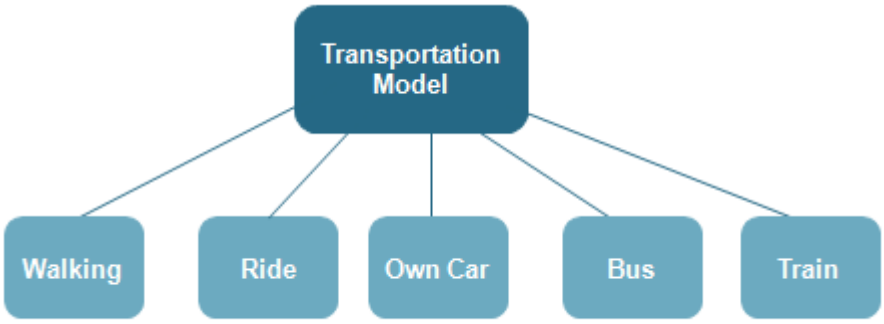


Figure 10 – Decision tree

Although, the transportation modals presented could be grouped in two sets following a price criteria: Walk and Ride are for free, so they will be grouped in a set named Free and Car, Bus and Train are going to be grouped in a set named Money. This new division shows that a new decision must be done before chose the transportation modal, the person needs to decide first whether to pay or not, in order to go to work and after that, the space of options is more restricted. The Figure 11 shows the decision tree based on this classification.

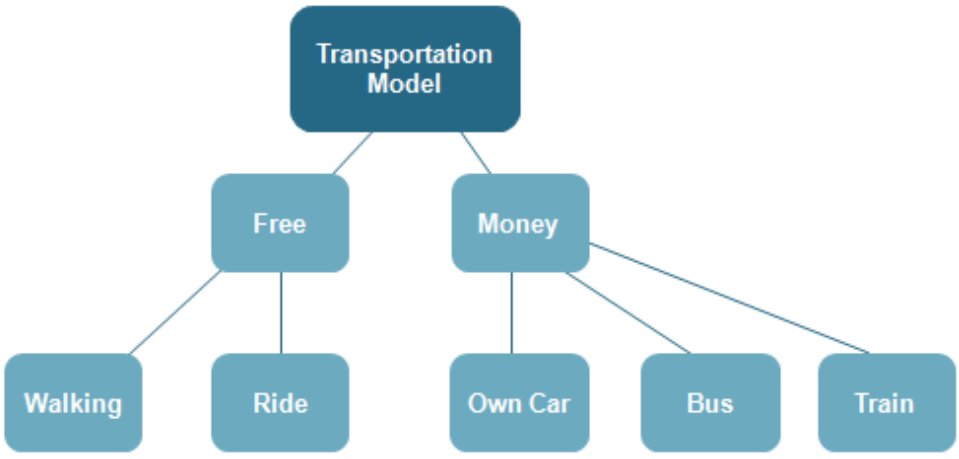


Figure 11 – Decision tree with new classification

This classification allow the resolution of this problem be modeled in a Nested Logit way, once the new groups will be the nests. Also, the elements in each nest are more similar among them in comparison with other elements from other nests.

4.2. Logit with unobserved characteristics

The Logit with unobserved characteristics is a different approach to obtain a Nested Logit model. It is also based on the consumer utility model. The Equation 11 show the utility model used to describe the choice of an alternative.

The model will be based in the creation of two markets named Inside Good and Outside Good, where the first represents the real market, it is what exits at that moment, and the second represents the potential market, it is the unexplored market, in other words is everything that one day can be achieved. The Logit with unobserved characteristics will be based on these two concepts and the base equation for the model will be developed by the next steps presented below.

$$u_{ij} = \delta_j + \varsigma_{ig} + \varepsilon_{ij}$$

Equation 11

The utility obtained by the consumer *i* due to the choice of the product *j*, in the set of products *g*, is described by the sum of the three terms: δ_j, which is the utility intrinsic to the choice, ς_{ig} which is a variable common to all products in set *g*, and ε_{ij} is the random term. Besides, the ς_{ig} depends on σ, which it is a term that shows the correlation between the sets, where 0 means no correlation and 1 means a complete correlation.

The modeling of δ_j is presented by the Equation 12, where the first term on the right side shows x'_j which is a vector that includes all the observable characteristics of the choice *j* with β being its coefficient. The second term ξ_j is a vector of non-observable characteristics and this is what also makes this option *j* different for different individuals.

$$\delta_j = x'_j\beta + \xi_j$$

Equation 12

D_G is a term defined to help in some of the following calculations. D_G is shown in the next equation, where N_g represents the nest *g*.

$$D_G = \sum_{j \in N_g} e^{\delta_j/(1-\sigma)}$$

Equation 13

So, Equation 14 shows the market share of the product *j* inside the group (nest) *g* as it was presented by Berry (1994). Considering that every nest has a group of products, the share presented by Equation 14 represents the force of the product *j* inside group *g* when compared to the other products inside *g*.

$$s_{j/g} = \frac{e^{\delta_j/(1-\sigma)}}{\sum_{j \in N_g} e^{\delta_j/(1-\sigma)}}$$

Equation 14

So, once D_G is defined previously, Equation 15 is just a substitution of the Equation 13 in the Equation 14.

$$s_{j/g} = \frac{e^{\delta_j/(1-\sigma)}}{D_g}$$

Equation 15

In the same way, Berry (1994) defined the share of a nest among the other nests as it is presented by the Equation 16. It presents the force of the products inside a share among the other shares. If a share has many products with strong share among all products in all nests, the share of the nest obviously will be also high.

$$s_g = \frac{D_g^{(1-\sigma)}}{\sum_g D_g^{(1-\sigma)}}$$

Equation 16

Thus, in order to obtain the share of a product among all nests, the following product was made. It considers the choice of the nest, represented by the S_g , and the the choice of the

product inside this nest, represented by $S_{j/g}$. So, the product of the two shares represents the share of the product among all nests as was said above.

$$s_j = s_{j/g} \cdot s_g = \frac{e^{\delta_j}}{D_g{}^{\sigma} \sum_g D_g{}^{(1-\sigma)}}$$

Equation 17

The probability associated with the outside good is going to be the only member of nest zero and is defined by Berry (1994) that $\delta_0 \equiv 0$ and $D_0 = 1$. So, Equation 18 shows the share associated with the outside good.

$$s_0 = \frac{1}{\sum_g D_g{}^{(1-\sigma)}}$$

Equation 18

Thus, $\ln(D_g)$ is given by the Equation 19 through the combination of Equation 17 and 18.

$$\ln(D_g) = \left[\ln(s_j) - \ln(s_0) \right] / (1-\sigma)$$

Equation 19

However, $\ln(D_g)$ is also obtained by the application of the natural logarithm on both sides of the Equation 15.

$$\ln(D_G) = \frac{\delta_j}{1-\sigma} - \ln(S_{j/g})$$

Equation 20

Using Equation 20 and 19,

$$\ln(s_j) - \ln(s_0) = \delta_j + \sigma \ln(S_{j/g})$$

Equation 21

Finally substituting the Equation 12 in the last equation, the final equation for the Logit with unobservable characteristics is shown in the Equation 22.

$$\ln(s_j) - \ln(s_0) = x'_j \beta + \sigma \ln(S_{j/g}) + \xi_j$$

Equation 22

5. The Problem

Brazil is a continental country divided in five different regions. Social and economical characteristics of each region are different. It is hard to create models to explain social and economical characteristics of Brazil due to its heterogeneity among the regions.

The air transportation is also quite different in each region. A unique model to describe the potential demand for air transportation is hard to be obtained with good accuracy due to differences of the states and regions.

The problem that is proposed in this work is to create a model to estimate the potential demand for regional air transportation in Brazil. This demand is supposed to considers the differences among the Brazilian regions. In order to include that, the problem will consider the country divided in microregions following the IBGE (Brazilian Institute of Geography and Statistics) division and for each one a potential demand will be made. So, it is going to be possible to evaluate if each Brazilian microregion has or not a potential for demand of regional air transportation.

A way to minimize eventual errors or an attempt to create a more believable model is to incorporate inside the model the differences among the regions. In the presented model**,** Oliveira and Salgado (2008) used some dummy variables to include the intrinsic characteristics of each region. Although, that is not the best option because it still necessary to insert some variables inside the model and this may cause some discrepancies in the fit of the curve to the data points.

A possibility to create a model with those characteristics is to use a Nested Logit model, as was described previously. The use of this model requires the creation of a decision tree with its nests and alternatives. In order to create a model with the differentiation of each region the Nested Logit seems to be a good alternative.

5.1. Modeling

The modeling of the problem was based in the Logit with unobserved characteristics. To evaluate the potential demand using this model it was necessary to define clearly how big is the actual market and how big the potential market is. In other words, the actual market is the size of the air transportation market in each of the microregions and the potential market is the market that was not achieved yet but it has possibility to be reached.

The size of the market was evaluated in terms of the number of seats used per microregion. Besides, the real number used to measure the size of the air transportation market in a microregion was a share, obtained through a ratio between the number of seats in the microregion and total seats in the Brazilian market.

The actual market and the potential market excluding the actual market represents the Inside Good and the Outside Good, respectively, as was said in the last topic about the Logit with unobserved characteristics.

The Inside Good represents the number of available seats for all available planes for the total number of flight hours for domestic commercial flights. Following the modeling presented in the last topic, each region has its own Inside Good as it is shown on Figure 12.

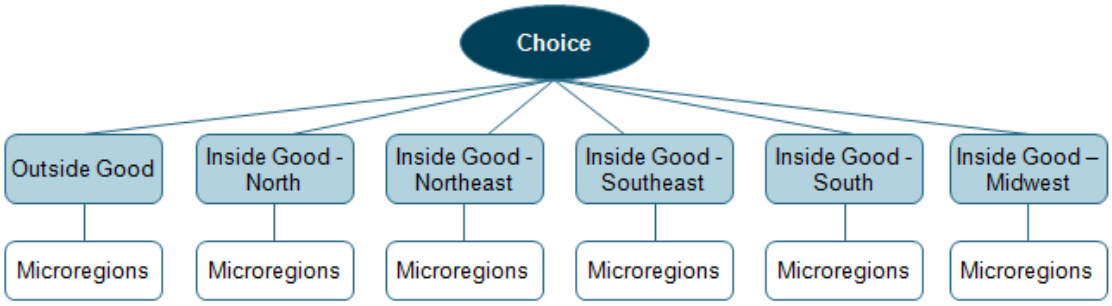


Figure 12 – Nested Logit modeling

The Outside Good was obtained through estimation of the average number of hours of domestic commercial flights in 2006, 2007 and 2008. The Table 4 shows the data extracted from the ANAC's information for total fleet, hours of flight and seats.

Table 4 – Total seats, Total hours of flight and total fleet from 2006 to 2008

Year	2006	2007	2008
Total seats	45.827.361	50.929.601	59.072.765
Total hours of flight	428.647	493.202	577.489
Total fleet	176	216	247

Then for each year the theoretical maximum number of hours of flight was estimated. Using the HOTRAN database, the higher number of hours of flight per day of an airplane was around 12 hours for the three years. This number was assumed to be the best hours of flight to obtain the maximum utilization of the airplane, being a hypothetical value. Thus, to obtain the maximum hours, the total fleet was multiplied by 12 hours and then by 365 days, what means one year. Then to estimate the maximum number of seats the ratio shown by the Figure 13 was made.

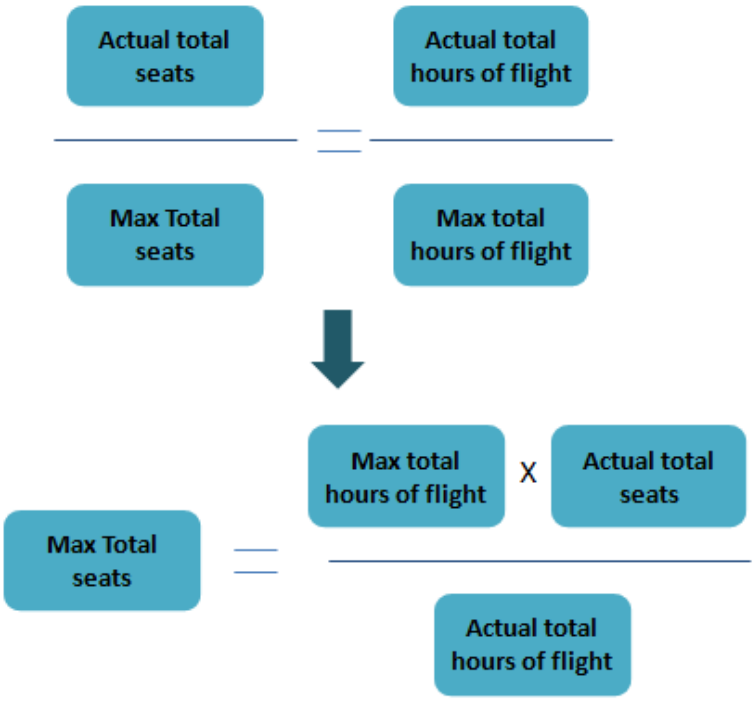


Figure 13 – Estimation of maximum number of total seats

So, the number of potential seats (Outside Good) was estimated with the following rational presented in the Figure 13. The results for Maximum and Total seats and Maximum total hours of flight are presented on the Table 5. From where it may be concluded that the Outside Good is approximately 85% higher than the Inside Good.

Table 5 – Maximum total hour of flight and maximum total seats from 2006 to 2008

Year	2006	2007	2008
Max total hours of flight	770.880	946.080	1.081.860
Max total Seats	82.415.965	97.695.261	110.666.130

With the values of Inside Good (Actual total seats) and Outside Good (Max total seats minus the Actual total seats) the model could be managed to go on.

The rational used to obtain the outside good was based only on the present situation of the air transportation in Brazil. In other words, the number of airplanes, the hours of flight, the number of available seats were all data from 2006 to 2008. In order to obtain a more accurate result for the inside good, an option would be use a more recent database.

Other possibility to obtain a better Outside Good would be add information about the capacity of the airports. Of course, it is not the most correct way to consider that all the airplanes will fly for 12 hours and the airports will have enough capacity to support all this demand.

Besides, the assumption used to calculate the Outside Good did not consider the fact that new airlines and new airplanes would enter in the Brazilian market. So, a more accurate modeling for this Outside Good should also involve these criteria.

5.2. Final Model

The Equation 21 shows the structure of the Nested Logit model based on the Logit with unobserved characteristics already described in the last topic. It has all the independent variables listed on the Table 6, a term related to the share of the microregion inside the region, which is the term that makes the link between the regions and makes, a random term that has the characteristic to incorporate all the factors not possible to the modeler to regard and include inside de other terms of the model.

Besides, the share of the outside good is also represented and the share of the microregion considering the total Brazilian market. The Equation 23 shown below is presented in a linear way with natural logarithms and in order to obtain the share of the microregion, s_j, is just to apply the exponential.

$$\ln(s_j) - \ln(s_0) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \sigma_1 \ln(s_{j/\text{region}}) + \varepsilon_j$$

Microregion Share | Outside Good Share | Independent Variables | Microregion Share Inside the Region | Random Term

Equation 23

The set of variables chosen to be part of the model is listed in the following Table 6

Table 6 – List of variables used in the model

Variable	Variable name for the model
GDP	sr_gdp
Boundary microregions GDP	xmr_gdp
Microregion area	sr_area
Distance of the closest airport	sr_kmnohb
Distance of the closest capital	km_med
Natural logarithm of the share of the microregion inside the region	Lnsjg
Constant	_cons

Two models were created with the variables presented on the Table 6. The first uses the database including all the microregions in Brazil, although only the microregions with domestic commercial flights were used to create the model. The amount of microregions considered was 131.

The model as created with all the variables listed for each microregion considered and is presented below. The column Coef shows the coefficients for each variable presented on Table 6. The variable presented by lnsjs0 represents the natural logarithm of the ratio between the share of the microregion and the share of the Outside Good.

The model that defines the existence of a potential demand for a determined Brazilian microregion is described in the following Figure 14.

Source	SS	df	MS
Model	689.200741	6	114.86679
Residual	39.8498186	124	.321369505
Total	729.050559	130	5.60808123

Number of obs	= 131
F(6, 124)	= 357.43
Prob > F	= 0.0000
R-squared	= 0.9453
Adj R-squared	= 0.9427
Root MSE	= .56689

lnsjs0	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sr_gdp	5.59e-09	1.74e-09	3.22	0.002	2.16e-09	9.03e-09
xmr_gdp	2.58e-09	1.40e-09	1.84	0.067	-1.88e-10	5.34e-09
sr_area	-3.81e-06	1.09e-06	-3.49	0.001	-5.96e-06	-1.65e-06
sr_kmmohb	-.0014571	.0003669	-3.97	0.000	-.0021833	-.0007308
km_med	-.0015088	.0006931	-2.18	0.031	-.0028805	-.000137
lnsjg	.8520918	.02676	31.84	0.000	.7991264	.9050573
_cons	-1.962328	.1991466	-9.85	0.000	-2.356495	-1.568161

Figure 14 – First model

The second model uses the same group of variables presented on the Table 6. Although, the database included only microregions which did not contain ae capitals with domestic commercial flights. The amount of microregions considered was 106.The model created is presented on the Figure 15 below.

```
. reg lnsjs0 sr_gdp xmr_gdp sr_area sr_kmnohb km_med lnsjg if capital==1

      Source |       SS       df       MS              Number of obs =     106
-------------+------------------------------           F(  6,    99) =  177.17
       Model |  360.316308     6  60.0527181            Prob > F      =  0.0000
    Residual |  33.5561397    99  .338950906           R-squared     =  0.9148
-------------+------------------------------           Adj R-squared =  0.9096
       Total |  393.872448   105  3.75116617           Root MSE      =  .58219

------------------------------------------------------------------------------
      lnsjs0 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sr_gdp |   1.64e-08   6.58e-09     2.50   0.014     3.37e-09    2.95e-08
     xmr_gdp |   2.79e-09   1.49e-09     1.88   0.063    -1.59e-10    5.74e-09
     sr_area |  -3.42e-06   1.16e-06    -2.94   0.004    -5.72e-06   -1.11e-06
   sr_kmnohb |  -.0014626   .0004331    -3.38   0.001    -.0023219   -.0006033
      km_med |  -.0008925   .0009652    -0.92   0.357    -.0028078    .0010227
       lnsjg |   .8459178   .0331815    25.49   0.000     .7800786    .9117569
       _cons |  -2.153226    .280439    -7.68   0.000    -2.709678   -1.596774
------------------------------------------------------------------------------
```

Figure 15 – Second model

Regarding the Equation 23, the reason to use only microregions with flights is because it is necessary to have a share for the microregion to input in the term $\ln(s_{j/region})$, otherwise the logarithm will be equals to zero and so, it will not be defined.

The two models obtained had a very good fit to the data, being the MacFadden's R^2 equals to 0,945 to the first model and 0,915 for the second model.

Besides, the share of all microregions where calculated by both models. For the microregions with actual flights, the models were applied exactly as they are presented before to calculate the shares. For the microregions with no flights, it is not possible to apply the model due to the existence of the term $\ln(s_{j/region})$ where the share cannot be zero. So, to evaluate the share of those cases with no flights, the models were applied using only the terms with coefficients $\beta's$ of the Equation 23. So, the term $\ln(s_{j/region})$ was equals zero.

That is good estimation for those cases, once the term $\ln(s_{j/region})$ represents the participation of a microregion inside its region and as it does not have any flights it is natural to imagine that the participation of it in the region is null. Of course, the best model possible will be one that do not need this kind of estimation, but this was necessary to have an estimation of the share.

Due to that, the comparison of the microregions will be limited. The share calculated through the models for those with actual flight will be compared only among them and in the same way the shares calculated for the microregions with no actual flights will be compared only among them, due to the differences in the equations.

Regarding the shares obtained for all the microregions with domestic commercial flights, the Table 7 shows a ranking of microregions with highest shares obtained through the use of the models. It is interesting to see that the top shares are located in state capitals, what is natural to expect due to the high level of development of those cities.

Table 7 – Top 10 Microregions with domestic commercial flights in 2008 by the two models

Top	Modelo with capitals	Modelo without capitals
1	São_Paulo	São_Paulo
2	Brasília	Rio_de_Janeiro
3	Rio_de_Janeiro	Brasília
4	Guarulhos	Guarulhos
5	Campanha_Central (Rio grande do Sul State)	Porto_Alegre
6	Porto_Alegre	Curitiba
7	Curitiba	Salvador
8	Salvador	Campanha_Central (Rio Grande do Sul State)
9	Manaus	Manaus
10	Recife	Belo_Horizonte

Now, regarding the shares obtained for all microregions without domestic commercial flights, Table 8 shows the ranking of the microregions with the highest estimated shares. The table also presents a column with the percent difference between the share obtained by the second model and the share obtained by the first model. It can be noticed that the microreegions with highest potential to have flights are located in the Southeast region and close to a huge city, as example of the top five which are near to São Paulo.

Besides, the result obtained by the second model presented a higher share in some cases, probably due to the absence of the capitals, the microregions with strong economy and near to a big city like São Paulo improved their importance among the other microregions.

The top 5 microregions presented by the Table 8 are located in São Paulo state, what really shows the force of São Paulo's economy compared to the other states. Osasco presented a high difference (57%) of shares obtained by the two models. For this case, the strong economy of the microregion should explain the high predicted share. The difference is

probably because of the absence of capitals which are the strongest economies in the second model. It is also interesting to see that a microregion is presented as a potential

Also, all the predicted shares for microregions without flights are from the region Southeast, what shows one more the importance of this region.

Table 8 - Top 10 Microregions without domestic commercial flights in 2008 by the two models

State	Model with capitals	State	Model without capitals	Diferença
SP	Osasco	SP	Osasco	57%
SP	Moji_das_Cruzes	SP	Santos	16%
SP	Itapecerica_da_Serra	SP	Moji_das_Cruzes	6%
SP	Franco_da_Rocha	SP	Itapecerica_da_Serra	4%
SP	Santos	SP	Franco_da_Rocha	-5%
RN	Litoral_Sul	RN	Litoral_Sul	-14%
RJ	Serrana	RJ	Serrana	-7%
RJ	Vassouras	RJ	Itaguaí	-8%
RJ	Itaguaí	RJ	Vassouras	-10%
RJ	Macacu-Caceribu	RJ	Macacu-Caceribu	-9%

Also, the Top 3 microregions without domestic commercial flights with highest shares per region estimated by the model without capitals are presented by the Table 9. This table shows the essence of the model, which is the use of it for identification of potential demand for air transportation in some cities, or location that do not have yet commercial flights. So, an interesting application of the model is also compare locations in the same region that do not have flight in a way to study which would have the stronger air transportation.

Table 9 – Top 3 microregions per regions predicted by the second model

Region	Microregion	Top
Mid-west	Anápolis	1
Mid-west	Rosário Oeste	2
Mid-west	Catalão	3
North	Castanhal	1
North	Rio Preto da Eva	2
North	Cametá	3
Northeast	Litoral Sul	1
Northeast	Suape	2
Northeast	Catu	3
South	São Jerônimo	1
South	Montenegro	2
South	Gramado-Canela	3
Southeast	Osasco	1
Southeast	Santos	2
Southeast	Mogi das Cruzes	3

The results presented until now are for a database created with the gross domestic product (GDP) of 2008 in all locations. So, another interesting analysis would be creating a different scenario for it.

Thus, it was considered the GDP growth per microregion on the years of 2006, 2007 and 2008. The average of these three growths was applied to the GDP of 2008 year over year until 2014 and, so, the value for it year was obtained per microregion. Then, the shares were calculated with the last model presented without the capitals.

After that, the top three highest shares per region were ranked and presented by the Table 10 below. It also includes the growth of the shares when compared with the share estimated by the same model using the GDP of 2008.

It is possible to notice that the highest share growth were obtained in the region southeast, probably due to the proximity of São Paulo city the GDP of them should be also high and model should have a high elasticity for the GDP. This estimation is good to show an application of the model in a future scenario.

Table 10 – Top microregions considering an estimated GDP for 2014

Region	Microregion	Top	Share Growth (2008 - 2014)
Mid-west	Anápolis	1	24.8%
Mid-west	Quirinópolis	2	3.8%
Mid-west	Entorno de Brasília	3	3.7%
Northeast	Valença	1	8.5%
Northeast	Macau	2	7.7%
Northeast	Feira de Santana	3	7.0%
North	Itacoatiara	1	2.3%
North	Ariquemes	2	1.5%
North	Castanhal	3	1.4%
South	Cruz_Alta	1	9.4%
South	Blumenau	2	9.2%
South	Paranaguá	3	8.1%
Southeast	Osasco	1	97.1%
Southeast	Jundiaí	2	40.4%
Southeast	Santos	3	32.8%

This prediction shown is an important example of how the model can be used in the future. The higher shares presented in the region Southeast region are probably due to the high GDP of them and also the higher GDP of the near microregions, once São Paulo city is a neighbor. Osascos presented the higher evolution until 2014 due to the high industrialization level experienced by the city in decade of 2000. So, as the growth of 2006 to 2008 was replicated to the following years until 2014.

Regions North presented the lowest evolutions of share until 2014. This was expected once it has a weak and primary economy based on the internal subsistence with a few industries. Besides, the distance from the center of the microregions to the closest airports area higher than in other regions due to its large dimensions and few number of airports if compared with Southeast, for example.

6. Conclusion

The use of discrete choice models for regional air transportation seems to be a good option for support in binary decisions. Probit and Logit are similar models that can generate closer results if used to generate models with same data.

It can be also concluded that both models created are consistent with its purpose. Both presented a good fit of curve with the data points and the variables chosen could explain different factors which can increase or decrease a demand for air transportation.

The second model generated, used a database without the Brazilian capitals, what give to the a model regional approach, once the most developed cities with the main airports were not used. This model can presented as the model proposed in the beginning of the work as the model for identification of potential in regional air transportation.

The equation of the model is shown previously. It can be applied to any microregion of Brazil, once the input data is given.

The predictions with the future scenario for the GDP in 2014 also show a good approach to predict future potential demands. It showed microregions without flight now days that can have future good demand for air transportation.

The model created can be worked further to improve its configuration or even to update the database that supports it. Besides, it is also possible to use it in different contexts trying to study groups of microregions or even new areas that would receive flights.

7. References

BERRY, Steven T., 1994. Estimating discrete-choice models of product differentiation. *Rand Journal of economics*,Vol. 25, Nº 2, pp. 242-262.

BIERLAIRE, Michel. Nested Logit Models. *Transport and Mobility Laboratory – Class notes.* Ecole Polytechnique Féderale de Lausanne

CIARLINNI, Marina, 2008. Modelos de escolha discreta e suas aplicação ao transporte aéreo, *Revista de Literatura dos Transporte, Vol 2, Nº 2, pp. 42-65*

LANGER, Wolfgang, 2000. The assessment of fit in the class of Logistic Regression Models: A pathway out of the jungle of pseudo-R^2s. Martin Luther University of Halle-Wittenberg,

NAGLER, Johnathan, 2001. Lecture Notes on discrete choice models. New York University

OLIVEIRA, Alessandro Vinícius Marques de, 2010. *Notas de Aula da disciplina Economia do Transporte Aéreo*. Instituto Tecnológico de Aeronáutica.

OLIVEIRA, Alessandro Vinícius Marques de. & SALGADO,Lúcia Helena, 2008. Constituição do marco regulatório para o mercado de aviação regional.

TRAIN, Kenneth, 2009. Discrete Choice Methods With Simulation. *Cambridge University Press, Chapters 3, 4, 5.*

VASISHT, A.K., Logit and Probit Analysis. *Indian Agricultural Statistics Research Institute.*

Brazilian Census 2008, IBGE (Brazilian Institute of Geography and Statistics).

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO	2. DATA	3. REGISTRO N°	4. N° DE PÁGINAS
TC	16 de novembro de 2011	DCTA/ITA/TC-080/2011	49

5. TÍTULO E SUBTÍTULO:

Application of a nested logit model for identification of potential demand in regional air transportation markets.

6. AUTOR(ES):

Paulo Thiago Araujo Moraes

7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES):

Instituto Tecnológico de Aeronáutica – ITA

8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR:

Air transportation; Discrete choice model; Nested logit.

9.PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO:

Transporte aéreo; Demanda (economia); Mercado; Modelos matemáticos; Administração de transportes; Transportes.

10. APRESENTAÇÃO: **X Nacional** **Internacional**

ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Alessandro Vinícius Marques de Oliveira. Publicado em 2011.

11. RESUMO:

Supported by a favored economical situation and its large dimensions, Brazil needs to include in studies that guides the public policies, issues of identification of potential demand for air transportation. The coverage of air transport throughout the country has dropped over the last decade, with a significant set of cities no longer served by regular aviation. The present work shows a model based on discrete choice models aimed at pointing to the potential of cities to be included in the meshes of regional airlines. A comparison between the traditional Probit and Logit models suggests that they are equivalent. A model Nested Logit with unobserved characteristics was used to rank the Brazilian microregions with potential sustainable economical operations. The results were presented considering the socio-economic scenario of 2008

12. GRAU DE SIGILO:

(X) OSTENSIVO **( ) RESERVADO** **( ) CONFIDENCIAL** **( ) SECRETO**