INSTITUTO TECNOLÓGICO DE AERONÁUTICA

Renan Rios Diniz

Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway

Trabalho de Graduação
2012

Civil-Aeronáutica

EVALUATION OF LONG-TERM DEMAND FOR THE VIRACOPOS AIRPORT UNDER ACCESSIBILITY HYPOTHESIS WITH OR WITHOUT HIGH-SPEED RAILWAY

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação

Renan Rios Diniz
Autor

Prof. Dr. Alessandro Vinícius Marques de Oliveira (Instituto Tecnológico de Aeronáutica)
Orientador

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

São José dos Campos, 23 de Novembro de 2012

CDU: 629

Renan Rios Diniz

Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway

Orientador
Prof. Dr. Alessandro Vinícius Marques de Oliveira (Instituto Tecnológico de Aeronáutica)

Engenharia Civil-Aeronáutica

SÃO JOSÉ DOS CAMPOS

INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2012

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

Diniz, Renan Rios
Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway/ Renan Rios Diniz.
São José dos Campos, ano.
Número de folhas no formato 103f.

Trabalho de Graduação – Engenharia Civil-Aeronáutica –
Instituto Tecnológico de Aeronáutica, 2012. Orientador: Prof. Dr. Alessandro Vinícius Marques de Oliveira.

1. Aeroportos. 2. Transporte de passageiros. 3. Transporte ferroviário. 4. Estimativas. 5. Planejamento de aeroportos. I. Instituto Tecnológico de Aeronáutica. II.Título

REFERÊNCIA BIBLIOGRÁFICA –

DINIZ, Renan Rios Diniz. **Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway**. 2012. 103 folhas. Trabalho de Conclusão de Curso. (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Renan Rios Diniz
TÍTULO DO TRABALHO: Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway
TIPO DO TRABALHO/ANO: Graduação /2012

Renan Rios Diniz
Rua José Krempser, Quadra U, Bloco 4, Olaria, Rio de Janeiro - RJ

I dedicate this work to God, in whose
love all things shall one day be
understood, and we shall see all not as a
mirror, but as they are, face-to-face.

ACKNOWLEDGEMENTS

Both graduating at ITA and finishing this graduation thesis would not have been possible without the aid of many people.

I want to say thank you to my teachers at ITA, who helped me both academically and personally. Among them I especially acknowledge Professors Maísa Oliveira and Paulo Ivo, my counselors at ITA, who helped me at many moments. I am also thankful to Professor Tânia Rabello, who was to me a model teacher both in the academic and human senses.

I am grateful to Professor Alessandro Oliveira, my advisor for this thesis, for showing me a world of possibilities in air transportation and econometrics, and for his patience during the research. I was also fortunate to have the valuable of opinions of Professor Carlos Müller as the reviewer for this thesis.

1st Lieutenant Mayara offered me much help for this work, directly and indirectly. Her notes also proved essential during my graduation.

I was also very happy to be supported by CNPQ, as part of this study was begun in their Scientific Introduction program.

A considerable part of this work was based on the study done by McKinsey and Company of the Brazilian airports, in partnership with FIPE and ITA. I cannot begin to understand the work that must have been put into this by consultants, researchers and teachers. I am fortunate to be able to give some contribution to such a large effort.

I am also in debt to many of my colleagues. I am especially grateful to by roommates at apartment 223, Gilberto, Armando, Paulo, Fábio and Renato. Among by friend I am also indebted to Gentil, Gabriel Grossi, Toledo and Walter, who are more important to me than they realize.

I thank also my parents, for giving me love and life.

I am also grateful to St. Anne and the Virgin Mary, and to God, without whom nothing would have been possible – thesis, university, people and life.

Can you fasten the harness of the Pleiades, or untie Orion's bands?

Can you guide the Crown season by season and show the Bear and its cubs which way to go?

Have you grasped the celestial laws? Could you make their writ run on the earth?

Can your voice carry as far as the clouds and make the pent-up waters do your bidding?

Will lightning flashes come at your command and answer, 'Here we are'?

Job 38, 31-35, New Jerusalem Bible

You gain strength, courage and confidence

by every experience in which you really stop to look fear in the face.

You are able to say to yourself,

“I have lived through this horror. I can take the next thing that comes along.”

You must do the thing you think you cannot do.

Eleanor Roosevelt –You Learn by Living

RESUMO

Este trabalho de graduação busca entender o impacto da construção ou não do TAV – um trem de alta velocidade ligando Rio de Janeiro, São Paulo e Campinas – nos aeroportos da Região de São Paulo, em particular o aeroporto de Viracopos, em Campinas. Isto é realizado por meio da definição de um modelo de demanda para os aeroportos das Regiões Metropolitanas de São Paulo e Campinas, e pelo desenvolvimento de modelo de escolha discreta os aeroportos de Viracopos, Guarulhos e Congonhas, nesta região. O modelo de demanda se baseia em uma revisão das estimativas realizadas pela empresa McKinsey (2009). As revisões incorporam um ajuste para mudanças na distribuição de renda com base em uma métrica desenvolvida por Sen (1982), além de ajustes no PIB, yield e no tráfego de passageiros em 2009. O modelo de escolha discreta desenvolvido para os aeroportos, embasado em uma pesquisa bibliográfica, é um modelo logit condicional segmentado entre as Regiões Metropolitanas de São Paulo e Campinas e o interior do Estado de São Paulo. Frequência e tempo de acesso se mostram relevantes no primeiro caso, e somente frequência no segundo, formando, em relação à influência da experiência dos passageiros com os aeroportos, um contraste com trabalho anterior desenvolvido por Moreno e Müller (2003). O modelo de escolha discreta é integrado à previsão de demanda e utilizado para discutir os impactos na utilização de cada aeroporto em 2030, que mostram possibilidade de rápida saturação dos aeroportos e a necessidade de melhoria do acesso a Viracopos. O trabalho conclui com recomendações de políticas e de pesquisa com base nos resultados.

Palavras-chave: Aeroporto de Viracopos, TAV, Previsão de demanda, Escolha discreta, Logit condicional

ABSTRACT

This graduation thesis seeks to understand the impact of building or not building the TAV – a high speed railway linking Rio de Janeiro, São Paulo and Campinas – on the São Paulo Area airports, especially the Viracopos airport, in Campinas. This is done through the definition of a demand forecast model for the airports the Campinas and São Paulo Metropolitan areas, and through the development of a discrete choice model for the Viracopos, Guarulhos and Congonhas airports, in the same region. The demand forecast model is based on a review of the estimates done by the company McKinsey (2009). Such reviews incorporate an adjustment for changes in income distribution based on a metric developed by Sen (1982), as well as adjustments to GDP, yield and the passenger traffic in 2009. The discrete choice model developed for the airports, supported by literature review, is a conditional logit model with segments for the São Paulo and Campinas Metropolitan Areas, taken together, and the rest of the São Paulo State. Frequency and access time are relevant variables in the first case, and only frequency in the second, creating, regarding the influence of passenger experience with airports with a previous work by Moreno and Müller (2003). The discrete choice model is integrated to the demand forecast and used to discuss the impacts on the use of each airport in 2030 showing a possibility of quick saturation of the airports and a need for improving access to Viracopos. The thesis concludes with policy and research recommendations based on the results.

Keywords: Viracopos Airport, TAV, Demand forecast, Discrete choice, Conditional logit

LIST OF ABBREVIATIONS AND ACRONYMS

ATL .... Hartsfield Jackson Atlanta International Airport/Atlanta
Arr .... Arrivals
ANAC ....Brazilian National Civil Aviation Agency (Agência Nacional de Aviação Civil)
BEL .... Val de Cans Airport/Belém
BSB ....Presidente Juscelino Kubitschek Airport/Brasília
BWI .... Baltimore-Washington International Airport/Washington
CDG .... Charles de Gaulle Airport/Paris
CGH ....Congonhas Airport/São Paulo
CNF .... Tancredo Neves Airport/Belo Horizonte
CWB .... Afonso Pena Airport/Curitiba
DAC ....Former Brazilian Department of Civil Aviation (Departamento de Aviação Civil)
DCA ....Washington Reagan International Airport/Washington
Dept .... Departures
DWF ....Direct Weekly Frequencies
FIPE .. Fundação Instituto de Pesquisas Econômicas (a Brazilian economic research institute)
FOR .... Pinto Martins Airport/Fortaleza
FLN ....Hercílio Cruz Airport/Florianópolis
GDP ....Gross Domestic Product
GIG ....Galeão Airport/Rio de Janeiro
GYN .... Santa Genoveva Airport/Goiânia
GRU .... Guarulhos Airport/São Paulo
HND .... Haneda Airport/Tokyo
HOTRAN ....Transport Time (Horário de Transporte)
IAD .... Washington-Dulles International Airport/Washington
IAC .... Former Brazilian Civil Aviation Institute (Instituto de Aviação Civil)
ID .... Identification
ITA .... Instituto Tecnológico de Aeronáutica
LAX .... Los Angeles International Airport/Los Angeles
LHR .... London Heathrow Airport/London
MAO .... Eduardo Gomes Airport/Manaus
METRO .... Metropolitan Area
NAT ....Augusto Severo Airport/Natal
OAK ....Oakland International Airport/Oakland
ORD ....Chicago O´Hare International Airport/Chicago
pax ....Passenger
PEK ....Beijing Capital International Airport/Beijing
PLU .... Pampulha Airport/Belo Horizonte
POA .... Salgado Filho Airport/Porto Alegre
REC .... Gurarapes Airport/Recife
SDU .... Santos Dumont Airport/Rio de Janeiro
SFO ....San Francisco International Airport/San Francisco
SJO ....San José International Airport/San José
SSA ....Eduardo Magalhães Airport/Salvador
Std. ....Standard
STS .... Sonoma County Airport/Sonoma

TAV ....Brazilian High Speed Train between Rio de Janeiro, São Paulo and Campinas (Trem de Alta Velocidade)
VCP .... Viracopos Airport/São Paulo
VIX..Eurico de Aguiar Salles Airport/Vitória

LIST OF SYMBOLS

Greek Symbols

$\boldsymbol{\beta_{0,i}}$	Intercept coefficient for choice i in conditional logit model or binomial logit model
$\boldsymbol{\beta_j}$	Coefficient for variable j in binomial logit model
$\boldsymbol{\beta i,_j}$	Coefficient for choice I and variable j in conditional logit model
$\boldsymbol{\varepsilon_{mj}}$	Random component of utility for consumer m for choice j
$\boldsymbol{\mu}$	Unadjusted income

Latin Symbols

atime	Access time variable in conditional logit model
dwf	Direct weekly frequencies variable in conditional logit model
exp	Experience variable in conditional logit model, given by number of times a route was taken by a passenger in the preceding 12 months
$\mathbf{f(\varepsilon_{mj})}$	Probability distribution function of random component of utility for consumer m for choice j
$\mathbf{F(\varepsilon_{mj})}$	Cumulative probability distribution function of random component of utility for consumer m for choice j
G	Gini index
L	Likelihood function for conditional logit model
$\mathbf{L(M_{intercept})}$	Calculated likelihood function for conditional logit model without explanatory variables
$\mathbf{L(M_{full})}$	Calculated likelihood function for conditional logit model with variables to be tested
$\mathbf{p_{ij}}$	Probability of choice of option j by agent i
$\mathbf{P\ (Y_i)}$	Probability of choice for choice i
$\mathbf{R^2}$	For logarithmic regression model, the goodness of fit measure, for conditional logit model, McFadden's R-squared
S	Inequality-adjusted income
$\mathbf{U_{mj}}$	Consumer utility of consumer m for choice j
$\mathbf{U_{mj}}$	Deterministic component of utility for consumer m for choice j
$\mathbf{y_{ij}}$	Binary choice indicator of choice j by agent i
$\mathbf{x_{i,j}}$	Variable i for choice j in conditional logit model
$\mathbf{x_i}$	Variable for choice i in binomial logit model
z	Statistic from normal distribution determined by standard deviation in estimated coefficients from conditional logit models

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Brazilian demand for air transportation has growth and is growing at a fast pace. This, along with the approach of major international events in Brazil, such as the 2014 FIFA World Cup and the 2016 Olympic Games, increasingly demands investments on expanding the Brazilian airports.

In this context the city of São Paulo and its surrounding presents a special challenge. São Paulo is the largest city in Brazil, with over 10 million residents, and is part of a metropolitan area inhabited by over 20 million individuals (IBGE, 2011b). Two airports directly serve São Paulo: the Congonhas Airport, in São Paulo proper, and the Guarulhos International Airport, in the neighboring city of Guarulhos.

A third airport, Viracopos, is located in the city of Campinas, 100 Km away from São Paulo. Viracopos fast growing airport with a considerable potential for expansion, and can be considered as part of the system of airports that caters to São Paulo. Currently access to Viracopos from São Paulo is done by two highways.

There is a project to build a high speed railway liking Campinas, São Paulo and Rio de Janeiro (the second largest city in Brazil), known as the TAV (acronym for Trem de Alta Velocidade – High Speed Train). Such a project, if actually built, might change considerably airport usage patterns in São Paulo.

Another factor in the problem is the Brazilian economy. In recent years Brazilian economy not just grew considerably (IBGE, 2012), but also reduced income inequalities, leading to changes in consumption patterns (Neri,2008).

Patterns in air transportation may also have changes. In 2011 and 2010, Brazilian air transportation grew at rates of over 10% (INFRAERO 2011, INFRAERO, 2010).

These questions make it very complex to plan investments on the São Paulo airports. Is Viracopos a feasible airport? How much should the airports expand if they are to offer a fair service level in an evolving economy? What are the factors that will determine demand for each airport?

This graduation thesis is not the first work to deal with these issues. A paper by Moreno and Müller (2003) studied airport choice in the São Paulo Metropolitan Area between the Guarulhos and Congonhas airports, and concluded that not only access time, but also flight frequency and the passenger's experience with each airport were important variables. Later, the Brazilian government commissioned a study by management consultancy McKinsey (2009), which developed demand projections for the Brazilian airports, albeit not dealing with differences resulting from variations at access time.

This graduation thesis seeks to expand on these studies. A review of the demand projections done by McKinsey is made, and also an adjustment of the model, including the incorporation of changes in the Brazilian income distribution through a metric developed by Sen (1982). The choice problem is also studied, and the study done by Moreno and Müller is expanded with the inclusion of the Viracopos airports into the choices, and the impact of frequency, access time and experience is assessed.

2 Purpose and structure of this work

2.1 Purpose

This study models long-term demand for the Guarulhos, Viracopos and Campinas airports for 2030, and uses these projections to recommend possible paths for evolving air transportation in the São Paulo area. This is done by means of a two-step approach: a review of projections for air transportation demand done by McKinsey (2009), at first at the national and then at the aggregate São Paulo area level, and then the development of a choice model between the three airports, using a discrete choice model.

Such a model takes into account access time, among other variables, making it possible to evaluate choice and demand in scenarios with or without the TAV, especially for the Viracopos airport.

2.2 Structure

Besides the Introduction and Purpose and Structure sections this study is organized into seven more sections.

In the third section, **Literature Review**, past works relevant to the subject matter are discussed.

In the fourth section, **Characterization of air transportation in the São Paulo Metropolitan Area**, an analysis of the airports in the São Paulo area and their characteristics is performed.

The fifth section, **Review of projections done by McKinsey (2009) and definition of demand baseline for airports in the São Paulo Area**, discusses McKinsey's demand projections and develops reviewed projections.

The sixth section, **Modeling of airport choice in the São Paulo Area**, is focused on building share division models for the airports in the São Paulo area based on logit modeling. This sections includes a Discussion on the Conditional Logit Model, a Description of Data Used, a Description of the Modeling Approach, and a discussion of the models.

The seventh section, **Analysis of access modes to the Viracopos Airport based on modeling** , leveraging on the models built, analyzes the impact of the TAV on forecasted demand

In the eight section, **Policy and Research recommendations**, comments are made on the results and their implications for public policy, as well as suggestions further study.

Finally in the ninth section, **Conclusions**, the findings of the study are summarized.

3 Literature Review

3.1 Review of papers modeling air transportation demand on multi-airport systems

Most research on airport choice models refer to a study by Skinner (1976, apud. Hess, 2006) as the first relevant attempt to model the choice behavior of passenger on multi-airport systems. In this study Skinner (1976) analyzed three airports in the Washington-Baltimore area: Washington National, Baltimore-Washington International and Washington Dulles International using a multinomial logit model.

The multinomial logit model had been recently developed as an econometric tool by McFadden (1973). This model makes it possible, in the context of discrete choice, given a sufficiently large database of passenger characteristics and choices, to forecast future choices based on determinate variables. The multinomial logit model is further explained in section 6.5.2.

There are two approaches to gather the needed data: asking subjects what their choices would be given certain conditions, that is, a declared preference approach, and researching actual choices given a set of parameters – a revealed preference approach

In the case of Skinner (1976) a revealed preference approach was adopted. The variables used were of two types: variables related to flight frequency at the chosen airport ad variables related the ease of access to the airport.

Regarding frequency, two direct flight frequency variables were used: one based on the number of weekday flights for the passenger's destination of choice and another based on the number of weekday flights for the passenger's destination and period of choice. As access variables were used the automobile access time and access utility.

The passengers were segmented into business and non-business groups, under an assumption that these groups might have different choice behaviors.

Although 16.000 interviews were conducted, only 1,552 questionnaires were actually used. Of these 918 were business travelers and 634 non-business.

Skinner (1976) concluded that all variables had statistical significance – that is – that they were relevant for choice. He also concluded that business travelers were more sensitive to accessibility and less sensitive to flight frequency. Also of note was his conclusion that only large changes in frequency were capable of changing significantly airport choice.

Another relevant paper was written by Harvey (1987), who studied the San Francisco Bay Area using a multinomial logit model. The San Francisco Bay Area is served by three airports: San Francisco International (SFO), Oakland International (OAK) and San José Municipal (SJC).

Harvey (1987) also adopted a revealed preference approach, using 2800 questionnaires, of which 1860 were valid.

Like Skinner (1976), Harvey (1987) used access and frequency variables. The access variable used was access time, and the frequency variables used were the number of frequencies to the chosen destination, considering connections, and the number of direct frequencies to the chosen destination.

Harvey (1987) concluded that direct flights have greater probability of choice than flights with connections, that frequencies, after a certain level, seem to reach a cap in their attractiveness ability, that the choice of access mode does not impact airport choice , that very high access times have a lower negative impact on attractiveness and that shorter flights were more time-sensitive.

Windle and Dresner (1995) provided a follow-up to the work done by Skinner (1976), again analyzing the Baltimore-Washington area using a multinomial logit model.

They also used a revealed preference approach. Windle and Dresner (1995) segmented their passengers into four groups: 948 residents travelling on business, 546 other residents, 1947 non-residents travelling on business and 1,041 other non-residents.

Windle and Skinner (1995) used, as others before access time and weekly flight frequencies to each chosen destination as variables. They added, however, a third type of variable related to experience with the airport.

Windle and Dresner (1995) reiterated the conclusion achieved by Skinner (1976): that business travelers were more time-sensitive. They, however, also found that they had increased sensitivity to frequency. It was also observed that non-residents were more sensitive to frequency, possibly due to lack of knowledge of the possibility of choice between airports.

Furthermore, the found out that experience was a very significant variable for choice, demonstrating the relevance of considering this factor in airport choice studies.

Pels, Nijkamp and Rietveld (1998) again analyzed airport choice in the San Francisco Bay Area, adding the Sonoma County airport to the three airports analyzed by Harvey (1987).

They again used a revealed preference approach, interviewing 21,459, and again segmented passengers into business and non-business categories. 5,016 business questionnaires were effectively used, and 6,249 other questionnaires were used.

Instead of using a multinomial logit approach they used a nested logit approach, which was then being used by other researchers in the field, such as Vovsha (1997) for the Tel-Aviv area (Hess, 2006a). The nested logit model elaborates on the multinomial logit model by considering that alternatives can be grouped on “nests” of related alternatives, therefore implying groups of choice order.

Pels, Nijkamp and Rietveld (1998) used as variables flight frequency, access time, number of seats per flight and air fares for each passenger.

The nested logit model allowed insights into choice order, obtained through different configurations of the models. Pels, Nijkamp and Rietveld (1998) concluded that airport choice precedes air carrier choice. They also concluded that seasonality affected considerably variable behavior. Interestingly they also concluded that business and non-business passengers had similar behaviors when seasonality was taken into account.

Moreno and Müller (2003) produced a paper of particular interest to this study as it dealt with the problem of airport choice in the São Paulo Metropolitan Area. Moreno and Müller considered choice between two airports: the Guarulhos International Airport (GRU) and the Congonhas Airport (CGH).

Moreno and Müller (2003), for their revealed preference approach, interviewed 1,923 passengers, 897 at GRU and 1,026 at CGH.

They used three kinds of variables: frequency variables, access time variables and experience variables.

They evaluated an array of frequency variables, including direct and indirect frequencies. The access time used was terrestrial access time, due the fact that most passengers used cars or equivalent transportation modes to reach airports. Access time was calculated based on a distance matrix for different neighborhoods as well as a matrix of average speeds along key arterials at different times. Experience was represented by the number of times each airport was used by passengers in the past year.

Moreno and Müller (2003) concluded that experience was the most important variable, followed by access time. They also found out that frequencies were more relevant for CGH than GRU.

Hess and Polak (2006b) modeled choice between the SFO, OAK and SJC airports in the San Francisco Bay area. They used as part of their revealed preference approach 5,091 of a set of over 21,000 questionnaires done in 1995 by the Metropolitan Transport Commission. Their approach used both nested logit and multinomial logit models, for comparison.

Hess and Polak (2006b) segmented passengers into in two manners: by residency/non-residency and by travel purpose: business, holiday or friend or family visits.

They used as variables frequency for each destination, access time, access cost, aircraft type (jet compared to turboprop) and experience, considering the number of flights taken each year by users at the three airports.

Hess and Polak (2006b) concluded that nested logit models were slightly superior to multinomial logit models, although both had very good fit. They observed as well that adding experience data to the models greatly improved results. They also highlighted the continued relevance of frequency and access time variables, while fares and aircraft sizes were important only for some segments. Finally they found positive results segmenting passengers into resident and non-resident categories.

Gelhausen and Christopher (2008) set out to add a new dimension to airport choice problems: adding a capacity constraint. In order to achieve this they add the concept of synthetic price. Synthetic price as defined by Gelhausen and Christopher (2008) is a function or airport use and capacity. Synthetic price is zero for unconstrained airports, and increases with overcrowding.

Gelhausen and Christopher (2008) applied the technique for airport choice in Germany, finding that the procedure changed significantly the outcome.

The final paper reviewed was written by Correia, Nyama and Nogueira (2011).

Correia, Nyama and Nogueira (2011) again modeled airport choice in the São Paulo Metropolitan Area, this time for three airports: GRU, CGH and the Viracopos airport (VCP). They also considered six hypothetical locations for a fourth airport, seeking to determine the best location for a new air field by maximizing demand.

Correia, Nyama and Nogueira (2011) used a gravitational model for demand, instead of the more usual logit model. The gravitational demand, which is considerably simpler, assumes demand to be a function of attractiveness (denoted by frequencies) and repulsion factions (denoted by access times).

Correia, Nyama and Nogueira (2011) divided the city of São Paulo into "population centers", corresponding to city districts: central, east, west, south and north. They also divided the São Paulo state into population centers.

The equation that expresses the gravitational model used by Correia, Nyama and Nogueira (2011) is:

$$E_{ij} = \frac{\frac{S_j}{T_{ij}^a}}{\sum_j \frac{S_j}{T_{ij}^a}} (1 - G) \tag{1}$$

In equation 1 Eij is the expected demand from population center i to airport j, T_{ij} the access time from the centroid of the population center to the airport, S_j is an attractiveness factor which takes into account service level and frequencies. The exponent a is a coefficient which was assumed to be 2.

Correia, Nyama and Nogueira (2011) concluded that access time was critical, and therefore the best locations for new airports had to be close to São Paulo.

3.2 Summary of literature review

Table 1 summarizes the studies analyzed.

Table 1 - Summary of literature review

Author and Year	Airports	Choice model	Sample size
Skinner (1976)	DCA, BWI, IAD	Multinomial Logit	1,552
Harvey (1987)	SFO, OAK, SJC	Multinomial Logit	1,867
Windle and Dresner (1995)	DCA, BWI, IAD	Multinomial Logit	4,482
Pels, Nijkamp and Ritfeld (1998)	SFO, OAK, SJC, STS	Nested Logit	11,265
Moreno and Müller (2003)	GRU, CGH	Multinomial Logit	1,923
Hess and Polak (2006)	SFO, OAK, SJC	Multinomial Logit and Nested Logit	5,091
Gelhausen and Christopher (2008)	...	Multinomial Logit with capacity constraint	Not specified
Correia, Nyama and Nogueira (2011)	GRU, CGH, VCP	Gravitational model	Not applicable

4 Characterization of commercial air transportation in the São Paulo Metropolitan Area

São Paulo is served by four airports: Guarulhos International (GRU), Congonhas (CGH), Viracopos (CGH) and Campo de Marte (MAE). Of these only the first three operate commercial flights.

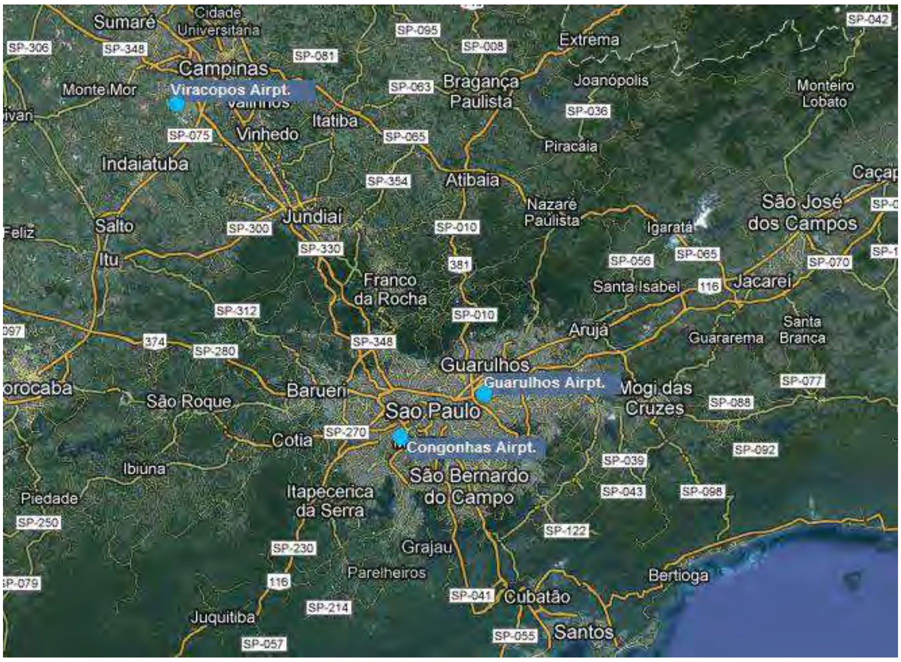

Figure 1 - Location of commercial airports in the São Paulo Metropolitan Area

Congonhas (Figure 2) is the oldest of the three airports, built in 1936. Congonhas has a very central location, and is 10 Km away from the traditional city center, besides being closer to the commercial centrality at the southern zone of the city. This central location, however, greatly restricts expansion possibilities for the airport.

Figure 2 - Congonhas Airport

Congonhas originally offered both domestic and international routes. After the Guarulhos International Airport was built in 1985, however, domestic flights predominated, an in 2008 the airport ceased to be considered international by the Brazilian National Civil Aviation Agency.

Congonhas, besides having no free area for expansion, features only two runways, and these cannot operate independently. Furthermore, the longest of these runways is only 1,940 meters long, too short for long-distance flights.

Due to capacity constraints Congonhas sees limited growth in transported passenger numbers. In 2009 Congonhas served 13,699,657 passengers, and in 2011, 16,753,591 passengers, according to INFRAERO.

The Viracopos International Airport (Figure 3) was homologated in 1960 in the city of Campinas, still inside the São Paulo State, but outside the São Paulo Metropolitan Area. Viracopos was originally designed to cater to long distance flights, with a runway 3,240 meters long. The airport's attractiveness for passengers from São Paulo, however, was diminished by the fact that a journey of over 95 Km was necessary to reach the airport from the São Paulo city center, especially after the Guarulhos International Airport was built.

Figure 3 - Viracopos Airport

The airport however, enjoyed greater success as a cargo airport. After 2008, when low cost company Azul Linhas Aéreas chose the airport as its São Paulo and national hub passenger traffic stated to grow at a fast pace.

Traffic, however, is still mostly restricted to domestic flights – as of August 2012 only 2 international routes operated at the airport. (INFRAERO 2012)

The airport has ample opportunities to expand its capacity, estimated by INFRAERO to be around 7.5 million passengers/year. Most of its terrain is unused, and its master plan foresees expansion for up to 90 million passengers/year (Brasil Econômico, 2012).

In spite of its distance from São Paulo access infrastructure. Two highways, which are not yet over capacity, link the Airport to São Paulo, and the project for the high-speed train linking Rio de Janeiro, São Paulo and Campinas contemplates a stop at the airport.

Viracopos was privatized in 2012, and its administration will be transferred from INFRAERO a private consortium, which will be also be responsible for expanding the airport's capacity.

In 2009 air passenger traffic was 3,221,161 passengers, and in 2011 7,568,384 passengers.

Guarulhos International (), in the city of Guarulhos, adjacent to São Paulo, is the newest of the airports, built in 1985, but quickly grew in importance to have the heaviest passenger traffic in the São Paulo area, and indeed, in Latin America. Guarulhos is located 30 Km away from the São Paulo city center.

Figure 4 - Guarulhos Airport

Guarulhos is equipped with two runways, one 3,700 meters long and the other 3,000 meters long. They cannot, however, operate independently. Due to urban occupation of the surrounding area it is not feasible to build a third runway (McKinsey 2009). Capacity can still grow, however, by terminal construction, and is forecast to grow from the current 31.4 million passengers/year to 55 million passengers/year.

In 2012 Guarulhos was also privatized, along the same policies set for the Viracopos airport.

In 2009 passenger traffic for Guarulhos International was 21,727,649 passengers, and in 2011 29,995,450 passengers.

Table 2 summarizes information about the three airports.

Table 2 - Characteristics of commercial airports in the São Paulo Area

Airport	Type of fligths	Distance from city center	Passenger Traffic (millions) - 2009	Passenger Traffic (millions) - 2011	Estimated Capacity (millions) - 2011	Potential Capacity (millions)
CGH	Domestic	10 km	13.7	16.8	~15	~15
GRU	Domestic and International	30 km	21.7	30.0	~30	~60
VCP	Domestic and International	95 km	3.4	7.6	~7.5	~90
Total	-	-	44.8	54.4	~52.5	165+

Sources: Traffic – INFRAERO (2012); capacities: INFRAERO (2012), McKinsey (2009) and Brasil Econômico (2012)

5 Review of projections done by McKinsey (2009) and definition of demand baseline for airports in the São Paulo Area

As stated in the Introduction to this study, it is the purpose of this article to determine potential long-term demand for the three commercial airports in the São Paulo Area considering changes in the mode of access to Viracopos. In order to do this it is necessary a) to forecast demand for the São Paulo Area; b) to develop a demand division model for the three airports.

As McKinsey (2009) developed a model for forecasting demand for the São Paulo airports.

It is the purpose of this chapter to evaluate such model and make necessary adjustments to recent developments, so that an first a national, and then a São-Paulo-specific estimate for air transportation demand is obtained.

5.1 Context for the demand projections according to McKinsey (2009)

In 2008 BNDES (Banco Nacional de Desenvolvimento Econômico e Social), in a scenario of growing concerns with overcrowding and lack of investments on the major Brazilian airports, hired McKinsey to study the Brazilian airports, estimate the demand for flights up to 2030 and give advice on investments to be done. Of particular concern was the investment plan for the São Paulo Area.

In the meantime the government continued to push for the construction the TAV linking Rio de Janeiro, São Paulo and Campinas, and therefore São Paulo with the Viracopos Airport. A study was done by a consortium between the Halcrow and Sinergia companies, purporting to show de projects, feasibility (Halcrow 2009).

McKinsey & Company advised the government to expand the Viracopos airport to its planned capacity, precluding the alternative approach of building a new airport. This solution raised a few questions.

First, as Table 3 shows, Viracopos is much farther from its main demand generation center than other major airports. As discussed literature review, access is one of the main variables for airport demand.

Table 3 - Distance from city center for the seven busiest airports worldwide and Viracopos

Airport	Road distance to City Center (km)
Atlanta (ATL)	11
Beijing (PEK)	32
Chicago (ORD)	27
London (LHR)	22
Tokyo Haneda (HND)	14
Los Angeles (LAX)	26
Paris (CDG)	25
Viracopos (VCP))	94

Second, the Brazilian government made successive public auctions for the TAV project, which failed. While a new auction is planned for 2012 the project's future became uncertain.

Third, as of 2011, McKinsey's demand projects had revealed themselves extremely conservative. Table 4 shows forecasts for 2014 done by McKinsey for the airports it studied in its most aggressive scenario compared to actual 2011 numbers. The 2011 traffic was already 96.4% of the 2014 aggressive forecast, with several cities already exceeding expected 2014 values. São Paulo was not an exception to the underestimation.

Table 4 - Comparison between McKinsey forecasts (aggressive scenario) and 2011 numbers

Airports	Aggressive McKinsey forecast for 2014	Actual 2011 numbers	2011 as percentage of 2014 forecasts
São Paulo aggregate	53,100,000	54,317,425	102.3%
Rio de Janeiro aggregate	24,200,000	23,466,423	97.0%
Belo Horizionte aggregate	8,900,000	10,328,291	116.0%
Brasília	17,900,000	15,398,737	86.0%
Curitiba	7,000,000	6,968,251	99.5%
Cuiabá	2,500,000	2,551,120	102.0%
Fortaleza	6,400,000	5,647,104	88.2%
Manaus	3,500,000	3,016,921	86.2%
Natal	2,800,000	2,586,220	92.4%
Porto Alegre	8,100,000	7,834,352	96.7%
Salvador	10,600,000	8,394,900	79.2%
Belém	3,400,000	2,995,547	88.1%
Florianópolis	2,900,000	3,122,035	107.7%
Goiânia	2,300,000	2,802,002	121.8%
Recife	7,800,000	6,351,249	81.4%
Vitória	3,500,000	3,182,394	90.9%
Total	164,900,000	158,962,971	96.4%

These issues make it necessary to develop updated estimations for air transportation demand in the São Paulo region and based on this discuss the validity of the TAV-based access solution to the Viracopos Airport.

In order to do this the first step is to better understand the model.

5.2 Examination of demand projections according to McKinsey (2009)

McKinsey (2009), developed two models in order to forecast air passenger demand for Brazilian airports, one of them using a bottom-up approach and another using a top-down approach.

The bottom-up model estimates demand for individual airports separately, based on existing routes. National demand is considered to be the sum of these projections. The top-down model estimates national demand first, and then uses estimates of airport traffic share for an interval of years to generate projections for each airport.

The bottom-up model is better suited to model particularities of each airport, but accounts neither for new routes nor changes in competitive relationships between airports, unlike the top-down model. The top-down model, therefore, is better suited for long-term forecasts, whereas the bottom-up methodology is precise for a short span of time.

As this study is concerned with long-term implications for the Viracopos Airport all subsequent analysis is based on the top-down model.

5.2.1 Model description

The top-down approach used as a reference the methodology used by the Brazilian Civil Aviation Institute (IAC) (McKinsey 2009).

The IAC used a logarithmic regression model based on air passenger traffic from 1979 to 2004 to estimate demand within a 30-year timeframe (IAC 2005). McKinsey, similarly, used data from 1979 up to 2008 to estimate demand for the years between 2010 and 2030. 2009 was extrapolated from partial yearly results and incorporated into the series, as shown on Table 5.

Table 5 – Estimated 2009 air passenger traffic for the Brazilian top 20 busiest airports (McKinsey 2009)

Airport	Passengers (2009 estimate)
Belém	2,193,018
Pampulha	581,916
Brasília	12,041,668
Confins	5,378,721
Curitiba	4,709,345
Cuiabá	1,633,797
Manaus	2,256,661
Florianópolis	1,943,137
Fortaleza	4,104,282
Galeão	11,255,153
Goiânia	1,594,865
Guarulhos	**21,064,330**
Campinas	**3,221,161**
Natal	1,815,758
Porto Alegre	5,431,298
Recife	5,073,329
Santos Dumont	5,249,149
Congonhas	**13,306,160**
Salvador	6,846,104
Vitória	2,357,823
Total Top 20	112,057,675

The logarithmic regression model used as explanatory variables national and global GDP, as well as national and international yield values. Additionally, a dummy variable was included accounting for the effects of the deregulation of the Brazilian air transportation market in 2002 (McKinsey 2009).

It was necessary to forecast GDP and yield values for the 2010-2030 timespan. National GDP was projected based on studies from the McKinsey Global Institute and MCM Consultores. Global GDP was forecast based on International Monetary Fund data, and the yield was predicted by McKinsey itself considering the influence of fuel prices, civil aviation competitiveness, productivity levels and the level of avoidable costs (McKinsey 2009).

Three scenarios were used – neutral, pessimistic and optimistic – each considering more or less aggressive assumptions for GDP growth and yield reduction. Table 6, Table 7, Table 8 and Table 9 show the values used.

Table 6 - Projections used by McKinsey for Brazilian GDP

Year	Pessimistic	Neutral	Optimistic
2009	-0.7%	-0.2%	0.3%
2010	3.4%	4.0%	4.8%
2011	3.7%	4.0%	4.6%
2012	3.6%	4.0%	4.6%
2013	3.5%	4.0%	4.7%
2014	3.0%	4.0%	5.0%
2015+	2.5%	3.5%	4.5%

Table 7 - Projections used by McKinsey for Global GDP

Year	Pessimistic	Neutral	Optimistic
2009	-2.7%	-2.1%	-1.6%
2010	0.5%	0.6%	0.8%
2011	2.1%	2.7%	3.4%
2012	2.3%	3.1%	3.8%
2013	2.4%	3.1%	3.9%
2014	2.3%	3.0%	3.7%
2015+	2.3%	3.0%	3.7%

Table 8 - Projections used by McKinsey for domestic yield

Scenario/Year	**2014**	**2020**	**2030**
Pessimistic	-7%	-8%	-13%
Neutral	-9%	-10%	-17%
Optimistic	-10%	-12%	-20%

Table 9 - Projections used by McKinsey for international yield

Scenario/Year	**2014**	**2020**	**2030**
Pessimistic	-2%	-3%	-4%
Neutral	-3%	-3%	-5%
Optimistic	-3%	-4%	-6%

Additionally McKinsey postulated a 5% increase in demand in 2014 and 2016 due to the FIFA World Cup in 2014 and the Olympic Games in 2016, hosted by Brazil. In this way it was possible to forecast the aggregate national air transportation demand.

In order to come to projections for individual airports it was necessary to forecast the share of each airport in the national network. This was done using 2008 market share values and then forecasting their evolution according to projections of regional GDP growth elaborated by McKinsey, as Table 10 and Table 11 show.

Table 10 – Domestic airport market shares based on passenger traffic (McKinsey 2009)

Airport/Area	Share 2008	Share 2009	Share 2014	Share 2020	Share 2030
METRO SP	29.17%	29.48%	28.34%	26.91%	24.43%
METRO RJ	13.60%	14.00%	13.97%	13.87%	13.68%
BSB	11.61%	12.02%	12.02%	12.04%	12.10%
SSA	6.56%	6.61%	6.85%	7.17%	7.74%
METRO BH	6.28%	5.67%	5.84%	6.06%	6.41%
REC	5.20%	4.95%	5.11%	5.31%	5.67%
POA	4.95%	5.19%	5.25%	5.32%	5.44%
CWB	4.92%	4.60%	4.68%	4.78%	4.94%
FOR	3.73%	3.99%	4.13%	4.31%	4.65%
BEL	2.47%	2.20%	2.28%	2.38%	2.57%
VIX	2.22%	2.36%	2.40%	2.46%	2.55%
FLN	2.22%	1.83%	1.86%	1.90%	1.95%
MAO	2.06%	2.15%	2.20%	2.27%	2.38%
GYN	1.76%	1.56%	1.60%	1.66%	1.76%
NAT	1.67%	1.73%	1.77%	1.83%	1.92%
CGB	1.58%	1.66%	1.69%	1.74%	1.81%
Sum	100.00%	100.00%	100.00%	100.00%	100.00%

Table 11 - Domestic airport market shares based on passenger traffic (McKinsey 2009)

Airport/Area	Share 2008	Share 2009	Share 2014	Share 2020	Share 2030
METRO SP	65.92%	64.32%	62.78%	62.78%	62.78%
METRO RJ	16.17%	19.87%	20.34%	20.34%	20.34%
BSB	1.29%	1.33%	1.42%	1.42%	1.42%
SSA	2.99%	2.68%	2.88%	2.88%	2.88%
METRO BH	1.25%	1.97%	2.13%	2.13%	2.13%
REC	1.69%	1.69%	1.80%	1.80%	1.80%
POA	3.80%	2.35%	2.48%	2.48%	2.48%
CWB	0.72%	0.66%	0.71%	0.71%	0.71%
FOR	1.80%	1.84%	1.96%	1.96%	1.96%
BEL	0.65%	0.28%	0.30%	0.30%	0.30%
VIX	0.00%	0.00%	0.00%	0.00%	0.00%
FLN	1.17%	0.76%	0.80%	0.80%	0.80%
MAO	1.23%	1.24%	1.31%	1.31%	1.31%
GYN	0.01%	0.01%	0.01%	0.01%	0.01%
NAT	1.31%	1.00%	1.07%	1.07%	1.07%
CGB	0.01%	0.01%	0.01%	0.01%	0.01%
Sum	100.00%	100.00%	100.00%	100.00%	100.00%

5.2.2 Model results

The model above described was calibrated by McKinsey, obtaining the coefficients for the variables shown on Table 12 and Table 13.

Table 12 - Econometric data for McKinsey's domestic air transportation demand model (McKinsey 2009)

Model variable	Description	Coefficient (Std. Error)	p-value
lnpibbr	Logarithm of Brazilian GDP	1.24234 (0.113004)	0.000
lnyieldd	Logarithm of domestic yield	-0.45439 (0.137297)	0.003
dlib	Dummy - deregulation	0.47542 (0.078995)	0.000
_cons	Constant	4.15246 (0.528803)	0.000

Table 13 - Econometric data for McKinsey's domestic air transportation demand model (McKinsey 2009)

Model variable	Description	Coefficient (Std. Error)	p-value
lngrav	Logarithm of global GDP	0.7771 (0.069576)	0.000
lnyieldi	Logarithm of international yield	-0.2795 (0.045209)	0.000
_cons	Constant	-0.8326 (0.391266)	0.043

Using these values, as well as the market shares from Table 10 . McKinsey obtained projections for individual airports for each year between 2010 and 2030.

Table 14 shows the projections for the neutral scenario.

Table 14 – Neutral projections for 2014, 2020 and 2030 air passenger traffic for the top 20 busiest airports in millions of passengers (McKinsey 2009)

	2014F	2020F	2030F	Yearly growth (%) 2009 -2030
Belém	3.1	4.1	7.1	5.7%
Pampulha	0.7	1.0	1.8	5.4%
Brasília	16.1	20.9	33.3	5.0%
Confins	7.3	9.8	16.4	5.5%
Curitiba	6.3	8.4	13.7	5.2%
Cuiabá	2.2	3.0	4.9	5.4%
Manaus	3.1	4.2	6.9	5.5%
Florianópolis	2.6	3.4	5.6	5.2%
Fortaleza	5.8	7.8	13.3	5.8%
Galeão	15.7	20.4	31.0	5.0%
Goiânia	2.1	2.8	4.8	5.3%
Guarulhos	**27.5**	**35.1**	**53.0**	**4.5%**
Campinas	**3.7**	**4.6**	**6.4**	**3.3%**
Natal	2.5	3.4	5.6	5.5%
Porto Alegre	7.3	9.7	15.7	5.2%
Recife	7.0	9.5	16.0	5.6%
Santos Dumont	6.2	8.0	13.6	4.6%
Congonhas	**16.9**	**20.6**	**30.3**	**4.0%**
Salvador	9.5	12.9	22.1	5.7%
Vitória	3.2	4.2	6.9	5.2%
Total T20	**149.1**	**193.9**	**308.4**	**4.9%**

Table 15 shows the projections for the pessimistic scenario.

Table 15 – Pessimistic projections for 2014, 2020 and 2030 air passenger traffic for the top 20 busiest airports in millions of passengers (McKinsey 2009)

Catchment area	2014F	2020F	2030F	Yearly growth (%) 2009 -2030
Metropolitan -São Paulo	46.0	53.6	70.0	3.0%
Metropolitan - Rio de Janeiro	20.9	25.2	34.9	3.7%
Brasília	15.4	18.6	26.1	3.8%
Salvador	9.1	11.5	17.3	4.5%
Metropolitan - Belo Horizonte	7.7	9.7	14.2	4.3%
Recife	6.7	8.4	12.6	4.4%
Porto alegre	7.0	8.6	12.3	4.0%
Curitiba	6.0	7.4	10.7	4.1%
Fortaleza	5.5	7.0	10.4	4.5%
Belém	2.9	3.7	5.5	4.5%
Vitória	3.0	3.7	5.4	4.1%
Florianópolis	2.5	3.1	4.4	4.1%
Manaus	3.0	3.7	5.4	4.2%
Goiânia	2.0	2.5	3.7	4.3%
Natal	2.4	3.0	4.4	4.2%
Cuiabá	2.1	2.6	3.9	4.2%
Total T20	**126.3**	**152.4**	**212.4**	**3.7%**

Table 16 shows the projections for the optimistic scenario.

Table 16 – Optimistic projections for 2014, 2020 and 2030 air passenger traffic for the top 20 busiest airports in millions of passengers (McKinsey 2009)

Catchment area	2014F	2020F	2030F	Yearly growth (%) 2009 -2030
Metropolitan -São Paulo	50.6	68.7	115.8	5.5%
Metropolitan - Rio de Janeiro	23.1	32.3	57.7	6.2%
Brasília	17.0	23.8	43.0	6.3%
Salvador	10.0	14.7	28.5	7.0%
Metropolitan - Belo Horizonte	8.5	12.3	23.4	6.8%
Recife	7.4	10.8	20.7	6.9%
Porto alegre	7.7	11.0	20.2	6.5%
Curitiba	6.6	9.5	17.6	6.6%
Fortaleza	6.1	8.9	17.2	7.0%
Belém	3.2	4.7	9.1	7.0%
Vitória	3.3	4.8	8.9	6.6%
Florianópolis	2.7	3.9	7.2	6.6%
Manaus	3.3	4.8	9.0	6.7%
Goiânia	2.2	3.2	6.1	6.8%
Natal	2.7	3.8	7.2	6.7%
Cuiabá	2.4	3.4	6.3	6.7%
Total T20	156.9	220.7	397.9	**6.2%**

5.2.3 Comparison to actual data

A way to test the model's effectiveness is to check whether the forecast air passenger traffic numbers match actual numbers already available.

These numbers can be obtained from INFRAERO, which has data for 2010 and 2011.

Table 17 compares forecast and actual numbers, considering the more optimistic estimation done by (McKinsey 2009).

Table 17 - Comparison of optimistic projections and actual statistics for passenger traffic in the main Brazilian airports

Year	Forecast	Actual	Error (%)
2010	119,259,811	140,725,886	-15.3%
2011	127,556,413	161,151,773	-20.8%

As Table 17 shows in 2010 the error was above 15%, and in 2011 surpassed 20%, possibly indicating a systematic downward bias. This bias, if true, accumulates over the forecast years, with grave implications for forecast quality by 2030.

Given the size of the error it is important either to adjust the model or adopt an alternate projection.

5.3 Adjustment of McKinsey model to actual results

5.3.1 Investigation on possible causes for underestimation

Even though the projections done by McKinsey proved themselves too conservative it might be possible to use them as a basis for discussion if the error sources are understood are eliminated or mitigated. Four sources of error were identified.

The first is the use of estimated, instead of actual, air passenger traffic statistics as the last year of the historical series, 2009. Table 18 shows a comparison of estimates and actual figures for 2009.

Table 18 - Comparison of McKinsey (2009) estimates and actual figures for air passenger traffic in the top Brazilian airports in 2009

Airport	2009 - McKinsey	2009 Actual	Relative error
Brasília	12,041,688	12,213,825	-1.4%
Campinas	3,221,161	3,364,404	-4.3%
Confins	5,378,721	5,617,171	-4.2%
Curitiba	4,709,345	4,853,733	-3.0%
Cuiabá	1,633,797	1,671,704	-2.3%
Fortaleza	4,104,282	4,211,651	-2.5%
Galeão-RJ	11,255,153	11,828,656	-4.8%
Guarulhos-SP	21,064,330	21,727,649	-3.1%
Manaus	2,256,661	2,300,022	-1.9%
Natal	1,815,758	1,894,113	-4.1%
Porto Alegre	5,431,298	5,607,703	-3.1%
Salvador	6,846,104	7,052,720	-2.9%
Belém	2,143,247	2,203,653	-2.7%
Pampulha	547,783	598,360	-8.5%
Florianópolis	1,982,882	2,108,383	-6.0%
Goiânia	1,472,962	1,772,424	-16.9%
Recife	4,694,445	5,250,565	-10.6%
Santos Dumont	4,831,780	5,099,643	-5.3%
Congonhas	14,150,092	13,699,657	3.3%
Vitória	2,183,144	2,342,283	-6.8%
Maceió	1,033,210	1,117,250	-7.5%
Total	**112,057,675**	**116,535,569**	**-3.3%**

It is possible to observe that the numbers were systematically underestimated, amounting to a total 3.3% lower than the actual.

A second source of error is the divergence between forecast GDP and yield values and actual values. Table 19 compares the national GDP envisioned by McKinsey (2009) and numbers provided by IBGE (2012).

Table 19 - Comparison of McKinsey (2009) estimates and actual figures for Brazilian GDP growth

Year	Pessimistic	Neutral	Optimistic	Actual
2009	-0.7%	-0.2%	0.3%	-0.2%
2010	3.4%	4.0%	4.8%	7.5%
2011	3.7%	4.0%	4.6%	2.7%

Source: IBGE (2012, 2011a, 2010)

Table 20 compares the variations in domestic yield envisioned by McKinsey (2009) and actual results according to data from ANAC (2012).

Table 20 - Comparison of McKinsey (2009) estimates and actual figures for Brazilian domestic yield

Year	Pessimistic	Neutral	Optimistic	Actual
2009	-1.4%	-1.8%	-2.1%	-20.8%
2010	-1.4%	-1.8%	-2.1%	-22.4%
2011	-1.4%	-1.8%	-2.1%	-4.6%

Source: ANAC (2012)

Table 21 compares the variations in international yield envisioned by McKinsey (2009) and actual results.

Table 21 - Comparison of McKinsey (2009) estimates and actual figures for international yield

Year	Pessimistic	Neutral	Optimistic	Actual
2009	-0.4%	-0.5%	-0.6%	-19.9%
2010	-0.4%	-0.5%	-0.6%	-4.9%
2011	-0.4%	-0.5%	-0.6%	n.a.

The analysis of GDP and yield data suggests that the demand surge in 2010 could have been caused by the fast GDP growth coupled with a steep price decline. The more moderate, but still high growth of 2011 could have been a consequence of still diminishing prices in a lower GDP growth economy.

A third possible source of error would be a change in the way the explanatory regression variables behave. The logarithmic regression model supposes that the relationship between demand and the logarithm of each explanatory variable is constant. As the series used goes back to 1979 it is reasonable to question whether the relationships for the more recent years and for the forecast future years are different from those at the beginning of the series. This might not be obvious in the model, as changes in a few years at the end of the series might not severely decrease the R^2 value,

A fourth source of error can possibly be the neglect of significant explanatory variables. This might involve not just the neglect of a variable relevant to the entire series, but also of a variable relevant only for the final years. In a manner similar to the above explained, this might not severely decrease R^2. This also might not be manifest in p-values and other measures of significance.

An important factor of change in the last few years was a remarkable reduction in income inequality in Brazil, coupled with relatively high GDP growth (Soares, 2008). This, according to researchers such as Neri (2008), has led to the formation of a new group of consumers dubbed as “The New Middle Class”. It is fair to assume that this might have an impact on the volume of air tickets bought.

These four presumed sources of error suggest each different adjustment approaches. Three are simple and direct: the substitution of projected 2009 passenger data for actual 2009 passenger data, the substitution of forecast GDP data for 2010 and 2011 for actual data, and the used of actual yield numbers. Results for these adjustments are shown in the next item. A separate item will be dedicated to the theoretical reasoning and the results for incorporating income inequality into the model.

5.3.2 Adjustment for actual passenger, GDP and yield data

The first, simpler step to improve the forecast model was to simply incorporate into the model actual data on demand, GDP and domestic and international yield from Table 18, Table 19, Table 20 and Table 21 respectively.

This yielded the results from Table 22.

Table 22 - Reviewed projections for the main Brazilian airports considering passenger, GDP and yield adjustments

Catchment areas	Airports	2014F	2020F	2030F	Av. Anual Growth (30F-09)
Metropolitan -São Paulo	GRU,CGH,VCP	54.7	74.7	125.3	5.9%
Metropolitan - Rio de Janeiro	SDU,GIG	25.1	35.4	63.0	6.6%
Brasília	Brasília	18.7	26.4	47.7	6.8%
Salvador	Salvador	11.0	16.3	31.4	7.5%
Metropolitan - Belo Horizonte	CNF,PLU	9.3	13.7	25.9	7.3%
Recife	Recife	8.1	12.0	22.9	7.4%
Porto alegre	Porto Alegre	8.5	12.2	22.3	7.0%
Curitiba	Curitiba	7.3	10.5	19.5	7.1%
Fortaleza	Fortaleza	6.7	9.8	19.0	7.5%
Belém	Belém	3.5	5.2	10.1	7.5%
Vitória	Vitória	3.7	5.3	9.9	7.1%
Florianópolis	Florianópolis	3.0	4.3	8.0	7.1%
Manaus	Manaus	3.6	5.3	9.9	7.2%
Goiânia	Goiânia	2.5	3.6	6.8	7.4%
Natal	Natal	2.9	4.2	8.0	7.2%
Cuiabá	Cuiabá	2.6	3.8	7.0	7.2%
Total Top-16		**171.2**	**242.7**	**436.6**	**6.7%**

It can be observed on Table 23 that, while numbers improved considerable error persists.

Table 23- Comparison of projections for air passenger demand from McKinsey, partially reviewed projections and actual figures

Year	McKinsey	Partially Reviewed projection	Actual	Error - McKinsey (%)	Error - Reviewed (%)
2010	119,259,811	128,749,297	140,725,886	-15.3%	-8.5%
2011	127,556,413	139,352,799	161,151,773	-20.8%	-13.5%

5.3.3 Procedure for adjustment for changes in income inequality

As previously said, the Brazilian economy has been characterized by a decrease in income inequality (Neri 2008, Soares 2008). The model developed by McKinsey, however, takes only GDP into account.

The Indian economist A. Sen (1982), points out that in the comparisons of the standard of living of nations based on GDP a "most serious difficulty is the treatment of income distribution". Sen, based on the idea of dispensing "with the time-honored device of drawing a distinction between the size and the distribution of national income" (Graaff, 1957, apud Sen 1982) proposes incorporating income and its distribution in a single metric.

One of the ways of expressing inequality is the Gini index. The Gini index a metric that varies between 0 (perfectly equal) and 1 (perfectly unequal), based on the Lorenz curve, that is, the curve that represents the cumulative distribution function of wealth (Gastwirth, 1971). The Gini index is given, in a probability versus income space, by the area between the Lorenz curve for a given population and a 45^o line, representing perfect distribution (Gastwirth 1972).

The unified metric proposed by Sen incorporates both income and the Gini index into a single metric, expressed by equation 2:

$$S = \mu(1 - G) \tag{2}$$

In equation 2 S is the adjusted income, μ is the unadjusted income and g is the Gini inequality index.

In the econometric model used by McKinsey GDP growth was used as parameter for defining the pessimistic, neutral and optimistic scenarios. A way of incorporating inequality into the model without needing to rebuild it is to consider GDP growth adjusted by Gini growth. Considering income and the Gini index as functions of time it is possible to derivate both sides and obtain:

$$\frac{dS(t)}{dt} = \frac{d\mu(t)}{dt} - \frac{d\mu(t)}{dt}G(t) - \mu(t)\frac{dG(t)}{dt} \tag{3}$$

Equation 3 gives the adjusted income growth as a function of income, the Gini index and their growth rates. That makes it possible to change data on the scenarios for GDP used by McKinsey to the values on Table 24:

Table 24 - GDP growth adjusted for inequality

Year	Adjusted Growth
2010	9.1%
2011	5.5%
2012	6.0%
2013	7.2%
2014	7.2%
2015+	5.5%

Gini index data were obtained from IPEA (Soares 2008, Barros 2009) and extrapolated for the subsequent years.

5.3.4 Results of adjustment for changes in income inequality

Table 25 shows the reviewed projections considering all adjustments.

Table 25 – Reviewed projections for the main Brazilian airports considering all adjustments

Catchment areas	Airports	2014F	2020F	2030F	Av. Anual Growth (30F-09)
Metropolitan -São Paulo	GRU,CGH,VCP	61.9	90.0	168.1	7.4%
Metropolitan - Rio de Janeiro	SDU,GIG	28.4	42.9	85.3	8.2%
Brasília	Brasília	21.3	32.4	65.7	8.4%
Salvador	Salvador	12.6	19.9	43.2	9.2%
Metropolitan - Belo Horizonte	CNF,PLU	10.6	16.7	35.6	9.0%
Recife	Recife	9.3	14.6	31.4	9.0%
Porto Alegre	Porto Alegre	9.7	14.9	30.6	8.6%
Curitiba	Curitiba	8.3	12.9	26.9	8.8%
Fortaleza	Fortaleza	7.6	12.0	26.1	9.1%
Belém	Belém	4.0	6.4	13.9	9.2%
Vitória	Vitória	4.2	6.5	13.6	8.8%
Florianópolis	Florianópolis	3.4	5.3	10.9	8.7%
Manaus	Manaus	4.1	6.4	13.5	8.9%
Goiânia	Goiânia	2.8	4.4	9.4	9.0%
Natal	Natal	3.3	5.2	10.9	8.9%
Cuiabá	Cuiabá	3.0	4.6	9.7	8.9%
Total Top-16		**194.6**	**295.2**	**595.0**	**8.3%**

Table 26 shows the effects on projections of the several adjustments performed.

Table 26 - Comparison of baseline and different adjusted projections

	2030 Estimate (millions)	Difference from baseline (%)
Baseline scenario	308.4	-
With 2009 data	322.4	4.,6%
Actual data + GDP adjustment + Yield adjustment	436.6	41.5%
Actual data + inequality adjustment + GDP adjustment + Yield adjustment	595.0	92.3%

It is possible to observe that the impact from 2009 data was small, and that GDP, yield and inequality adjustments were very meaningful.

5.3.5 Comparison to actual data

Table 27 compares the projections done by McKinsey (2009), the adjusted projections and actual numbers.

Table 27 - Comparison of projections for air passenger demand from McKinsey, fully reviewed projections and actual figures

Year	McKinsey	Reviewed projection	Actual	Error - McKinsey (%)	Error - Reviewed (%)
2010	119.259.811	135.107.900	140.725.886	-15,3%	-4,0%
2011	127.556.413	147.789.993	161.151.773	-20,8%	-8,3%

It is possible to see that the reviewed projections considerably reduce errors, even though they not achieve a perfect match.

5.4 Definition of final demand baseline for São Paulo area

The adjusted McKinsey model makes it possible to define a new demand baseline for the São Paulo area. Considering market shares as shown on Table 10 the values on Table 28 are obtained for the aggregate demand of the São Paulo area.

Table 28 - Aggregate demand for the São Paulo Area airports – Neutral baseline

Year	National Demand	São Paulo Domestic Share	São Paulo International Share	São Paulo Total Share	São Paulo Aggregate Demand
2014	194,599,845	28.34%	62.78%	31.79%	61,870,009
2020	295,190,443	36.91%	62.78%	30.51%	90,049,474
2030	594,996,829	24.43%	62.78%	28.24%	168,074,224

This demand forecast will be called from now on the neutral baseline.

A comparison of the neutral baseline with Table 2 indicates that by 2030 the airports will be exceed their maximum capacity of about 165 million passengers, already taking into account foreseen infrastructure improvements.

There are criticisms which can be done to the neutral baseline which might actually be optimistic. The first is that it assumes that income inequality will keep decreasing over time at a rate of 0.006 Gini points per year. A study done by IPEA (2007) indicates both that this is a high level of inequality reduction, compared to international benchmarks, and that Brazil had not experienced, until the current cycle, a period of decrease in income inequality longer than seven years since the Gini index began to calculated.

The second is a criticism to decreases in yield. It is possible that, as yields fell considerably in 2009, 2010 and 2011, that much of the potential for yield reduction is spent.

The third is the use of two high-growth years to test the effectiveness of the adjusted model. These high growth years might prove themselves to be exceptions to the model, and therefore poor guides for evaluating it effectiveness.

For these reasons it is interesting, in all analyses, to consider a more pessimistic scenario for forecasts. A possibility would be using McKinsey's average scenario as a forecast. However, such a scenario would probably be too pessimist, in the light of recent data discussed by this study.

A compromise can be achieved by averaging the two scenarios. Such numbers, for São Paulo, are shown on Table 29. This scenario will be henceforth called the conservative scenario, and also used for defining demands for the São Paulo Area airports.

Table 29 – Conservative baseline for the São Paulo Area airports

Year	São Paulo Aggregate Demand
2014	61,870,009
2020	90,049,474
2030	168,074,224

6 Modeling of airport choice in the São Paulo Area

With a demand baseline established the second step of this study is to develop an airport choice model for the São Paulo Area.

6.1 Modeling approach – the multinomial conditional logit model

The problem of choosing an airport in a region, as is the case for this study, is one that involves choice between non-quantitative, non-ordered options. The approach used so far - multinomial regression analysis, works well for quantitative phenomena. This new problem, however, demands a different approach.

McFadden (1973) recognizes this need, observing, furthermore, that the factors that influence choice are often outside the sphere of control of the researcher, and therefore demand a statistical approach based on sets of individual choices. He also sees that such an analysis may wield explanatory variables, but also be subject to an element of undetermined randomness. Finally, McFadden (1973) also observes that choice variables are not always attributes of the choosing agent, but also of the choices themselves. In face of this he proposes the multinomial conditional logit model or simply the conditional logit model.

The conditional logit model is an extension of the simpler binomial logit model. The binomial logit model seeks to predict individual choice behavior between two alternatives, based on a set of predictive variables.

The binomial logit model assumes that individuals seek to maximize utility trough choice. Mathematically, this is expressed by equation 4 (Cushing and Cushing, 2007):

$$U_{mj} = V_{mj} + \varepsilon_{mj} \quad (4)$$

Equation 4 shows that each choice j made by an agent n has associated to itself an utility U. This utility is defined by a set of variables (V) and by a random component **ε**. In the case of the logit model this component follows a Gumbel probability distribution. Such being the case, it can be shown (Bierens, 2008) that the probability of choosing option j over option k is given by:

$$P_{mj} = \frac{1}{1+e^{-\beta_0-\beta_1 x_1-\dots-\beta_n x_n}} \quad (5)$$

Equation 4 shows that the probability P for each agent m of choosing option j over k is determined by an exponential function with contains coefficients β and variables x.

In order to model choices between several options, however, it is necessary to extend this model.

A way of doing this is modeling the problem of choosing between alternatives from a set Y={1,2,..,K} is to model as a set of N-1 binomial logistic regressions in which one choice is used as a reference to which the probability of each other choice is compared.

The form of the utility function for each consumer is still given as in equation 3. In the case of the conditional logit model, it is assumed that the random component of utility follows the extreme value distribution, with probability density function in the form of equation 6 and cumulative density function in the form of equation 7 (Cushing and Cushing, 2007):

$$f\left(\varepsilon_{mj}\right) = e^{-\varepsilon_{mj}} e^{-e^{-\varepsilon_{mj}}} \quad (6)$$

$$F\left(\varepsilon_{mj}\right) = e^{-e^{-\varepsilon_{mj}}} \quad (7)$$

This being true, N-1 expressions in the form of equations 8 through 10 which, correspond to the probability of choosing case Y_i:

$$\ln \frac{P(Y_i=1)}{P(Y_i=N)} = \beta_{1,0} - \beta_1 x_{1,1} - \ldots - \beta_n x_{1,n} \quad (8)$$

$$\ln \frac{P(Y_i=2)}{P(Y_i=N)} = \beta_{2,0} - \beta_1 x_{2,1} - \cdots - \beta_n x_{2,n} \quad (9)$$

...

$$\ln \frac{P(Y_i=N-1)}{P(Y_i=N)} = \beta_{n-1,0} - \beta_1 x_{n-1,1} - \ldots - \beta_n x_{n-1,n} \quad (10)$$

Each choice Y_i has its own coefficients β_{ij}, where i is the choice and j represents the index for each variable.

Considering that the sum of all probabilities must be 1 it is obtained for choice N:

$$P(Y_i = N) = \frac{1}{1+e^{-\beta_{1,0}-\beta_1 x_{n,1}-\ldots-\beta_n x_{n,n}}} \quad (11)$$

Therefore it can derived for the other choices that:

$$P(Y_i = 1) = \frac{e^{-\beta_{n,0}-\beta_{n,1}x_{n,1}-\dots-\beta_{n,n}x_{n,n}}}{1 + e^{-\beta_{1,0}-\beta_{1,1}x_{1,1}-\dots-\beta_{1,n}x_{1,n}}} \tag{12}$$

$$(Y_i = 2) = \frac{e^{-\beta_{n,0}-\beta_{n,1}x_{n,1}-\cdots-\beta_{n,n}x_{n,n}}}{1 + e^{-\beta_{2,0}-\beta_{1}x_{2,1}-\cdots-\beta_{n}x_{2,n}}} \tag{13}$$

$$\dots$$

$$(Y_i = N - 1) = \frac{e^{-\beta_{n,0}-\beta_{1}x_{n,1}-\dots-\beta_{n}x_{n,n}}}{1 + e^{-\beta_{2,0}-\beta_{1}x_{2,1}-\dots-\beta_{n}x_{2,n}}} \tag{14}$$

If the variables in the model are only attributes of the individual the model is called a multinomial logit; if they are attributes of the choice the model is called conditional logit.

Such derivation required one important assumption to be made, so that equations 8-10 would be true: that the odds of choice between the choices in the set depend only on choices in the set. An implication of this is that removing or adding choices to the set does not change the relative probabilities between pairs of sets. This axiom, known as Independence of Irrelevant Alternatives, may not verify for all problems, and therefore must be taken into account if it is the purpose of the researcher to add or remove alternatives from his model.

Further discussion of the conditional logit model, including complete derivation from axioms and demonstrations can be found at (McFadden, 1973). A simpler approach can be found at (Cushing and Cushing, 2007)

6.2 Considerations on the estimation procedure for a conditional logit model

Using the logit model to estimate a given choice requires defining explanatory variables, estimating coefficients for each variable, evaluating the explanatory power of the model and verifying the whether the chosen variables are significant.

The first step depends on several factors. First, it is necessary to have a good qualitative understanding of the problem being modeled, so that the model can predict well results from these variables. Second, it is important to ascertain that a considerable number of observations is obtainable in order to avoid statistical noise. Third, the variables chosen are not correlated. A final important issue is to avoid overfitting – that is – using a model with many variables that is very capable of explain historic data, but is too rigid and unresponsive to change to explain the future (Stone and Rasp, 1993).

The second step is to estimate the confidents. That can be done using a maximum likelihood approach based on a logarithmic likelihood function, based on a data set. Such function is maximized via an iterative approach, such as the Newton-Raphson method, thus yielding values for the coefficients. (McFadden, 1973). Equation 15 shows the logarithmic likelihood function (Hoffman and Duncan, 1988):

$$\ln L = \sum_{i=1} \sum_{j=1} y_{ij} p_{ij} \tag{15}$$

In equation 15 y_{ij} is equal to 1, if agent i chooses choice j, or 0 if not. The parameter p_{ij} is the probability of choice of option j by agent i.

The third step is to evaluate whether the model as a good explanatory power. This is not as straightforward as it would be for linear regression, case in which R^2 provides a direct measure for goodness of fit. In the case of multinomial logistic regression McFadden (1973) proposes an alternative metric, henceforth called McFadden's R^2 or pseudo- R^2, given by equation 16:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} \tag{16}$$

In equation 16 L is the likelihood defined on equation 15. M represents the logistic regression model as calculated. M represents the model as it would be if it had no explanatory variables and therefore only an intercept coefficient.

McFadden's R^2 represents the degree to which a model has more explanatory power than a model with no variables. It cannot be said, therefore, that it directly measures goodness of fit. Nevertheless, it is still true that values vary between 0 and 1 and that higher values denote more powerful values, like common R^2. Good values for R^2, however, tend to be not as high – values in the range of 0.2-0.4 are considered satisfactory.

Finally, it is necessary to verify the significance of the variables. A variable is said to be significant if its coefficient is not null. The maximum likelihood estimation yields coefficients accompanied by standard deviations. This makes it possible to set up a range of values for the coefficients, within a given confidence value. It is usual to ensure 95% or 90% of confidence on the non-nullity of coefficients. Dividing each coefficient by its standard deviation a test statistic named z is obtained. This z statistic follows a normal distribution. This z value is then used as limit for a confidence test of whether the coefficient is not null. If the probability is greater than 5% or 10% (depending on the confidence level desired) it is not possible to reject the null hypothesis. This probability is called a p-value, denoted by $p>|z|$.

It is important to check whether the variables are not null, as if they were this means not only that the variable might lack actual predictive power, but also that its real world behavior might be opposite to the predicted.

6.3 Software used

It is necessary to use computational aid in order to deal with the large amounts of date and the iterations the multinomial logit approach requires. All multinomial logit analysis done in this study was done with the aid of the program Stata for Windows (32-bit), a software produced by StataCorp LP.

6.4 Description of data source

As part of the study done by McKinsey for BNDES a survey was made of air passengers on key Brazilian airports. This survey was carried out by FIPE – an economics research foundation linked to the University of São Paulo.

The survey was conducted in 30 airports – the 20 busiest airports in the country, plus an airport at each state capital and the Pampulha (Belo Horizonte – MG) and Porto Seguro (BA) airports, deemed strategic (FIPE, 2009). The survey set out to determine the cities which comprised the catchment area of each airport, the passenger’s profiles and the routes they took.

Of particular interest to this study, the survey collected data on access modes, income, door-to-door airport access time and passenger origin, down to the neighborhood level (FIPE, 2009).

The sample size at each airport was fixed having in mind a confidence rate of 95%, further adjusted considering small sample sizes at the smaller airports. (FIPE, 2009).

Table 30 shows the number of passengers interviewed at each airport.

Table 30 - Passengers interviewed by FIPE (2009) at Brazilian Airports

Aeroporto	Cidade - UF	Amostra Efetiva
Guarulhos - Governador André Franco Montoro	São Paulo - SP	7.593
Congonhas	São Paulo - SP	5.106
Galeão - Antônio Carlos Jobim	Rio de Janeiro - RJ	4.243
Pres. Juscelino Kubitschek	Brasília - DF	3.886
Deputado Luís Eduardo Magalhães	Salvador - BA	2.348
Tancredo Neves	Belo Horizonte - MG	1.982
Salgado Filho	Porto Alegre - RS	2.066
Guararapes - Gilberto Freyre	Recife - PE	1.798
Afonso Pena	Curitiba - PR	1.713
Santos Dumont	Rio de Janeiro - RJ	1.765
Pinto Martins	Fortaleza - CE	1.378
Val de Cans	Belém - PA	910
Hercílio Luz	Florianópolis - SC	725
Eduardo Gomes	Manaus - AM	745
Eurico de Aguiar Salles	Vitória - ES	830
Augusto Severo	Natal - RN	621
Santa Genoveva	Goiânia - GO	594
Marechal Rondon	Cuiabá - MT	566
Viracopos	Campinas - SP	477
Zumbi dos Palmares	Maceió - AL	387
Marechal Cunha Machado	São Luís - MA	409
Campo Grande	Campo Grande - MS	451
Santa Maria	Aracaju - SE	416
Pampulha - Carlos Drummond de Andrade	Belo Horizonte - MG	385
Macapá	Macapá - AP	383
Senador Petrônio Portella	Teresina - PI	394
Pres. Castro Pinto	João Pessoa - PB	384
Governador Jorge Teixeira de Oliveira	Porto Velho - RO	384
Presidente Médici	Rio Branco - AC	430
Brigadeiro Lysias Rodrigues	Palmas - TO	399
Boa Vista	Boa Vista - RR	402
Porto Seguro [3]	Porto Seguro - BA	388
Total		44.558

[1] Movimento Operacional Acumulado da INFRAERO/SINART - 2008 - Passageiros - embarque mais des

[2] No banco de microdados corresponde a Variável - Peso_embarques_totais

[3] Dados Sinart - Passageiros - embarque mais desembarque (internacional e doméstico)

6.5 Adjustment of data for modeling

As this the aim of this work is to specifically analyze the Guarulhos, Congonhas and Viracopos airports it was necessary to trim down the database so that it contained only the necessary data.

The first step was deleting surveys not done with passengers departing from GRU, VCP or CGH. After this filter was applied a total of 10,228 observations remained.

A second step was restricting the analysis to passengers departing from the São Paulo state. This second filter left a total of 9,200 observations, distributed between airports in the manner shown by Table 31:

Table 31 - Observations in data sample

Airport	Observations (absolute)	Observations (relative)
Guarulhos	4,657	50,62%
Congonhas	4,107	44,64%
Viracopos	436	4,74%
Total	**9,200**	**100,00%**

As discussed during literature review, frequency variables are of capital importance for airport choice models. A third step, therefore, was to incorporate frequency data into the data set.

In order to this it was necessary to obtain frequency data for the period in which the survey was made – July and August – 2009. This information is given by the HOTRAN – a daily compilation of registered flights at each airport done by the Brazilian national civil aviation authority (ANAC, 2009).

The HOTRAN includes regular flights of all types, including cargo. The HOTRAN also includes some flights, such as air bridge routes between Rio de Janeiro, São Paulo and Belo Horizonte as "special flights". As this study is only interested on commercial passenger flights only international, national, regional and special flights were considered for the HOTRAN released on 08/29/2009.

As some flights do not operate everyday, especially at Congonhas, it was decided to use weekly (direct) frequencies as the frequency parameter.

The frequency indicator varies for each passenger and airport, as each airport has a different number of flights for a chosen destination. The database, therefore, crosses passenger choice data in order to define for each passenger the frequencies for the selected flight for each airport.

Appendix B provides a list of the weekly frequencies from the three airports of interest to all regular commercial destinations.

Another important variable is access time. The survey conducted by FIPE (2009) included data on time spent by each passenger to reach the airport. Like frequency, however, it is important to know the access time to the airports not chosen. As it was not possible to obtain distance data it was necessary to approximate these distances.

In order to do this, passengers were segmented into geographical zones. Of these zones 8 were divisions within the São Paulo municipality (Figure 5), 3 were divisions within the outer São Paulo Metropolitan Area (Figure 6) and the two last ones were the Campinas Metropolitan Zone and the remnant of the São Paulo State (Figure 7).

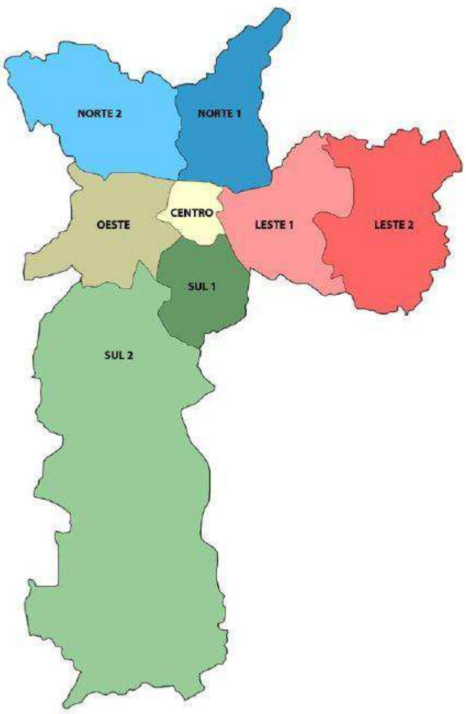

Figure 5 - Division of São Paulo municipality into zones

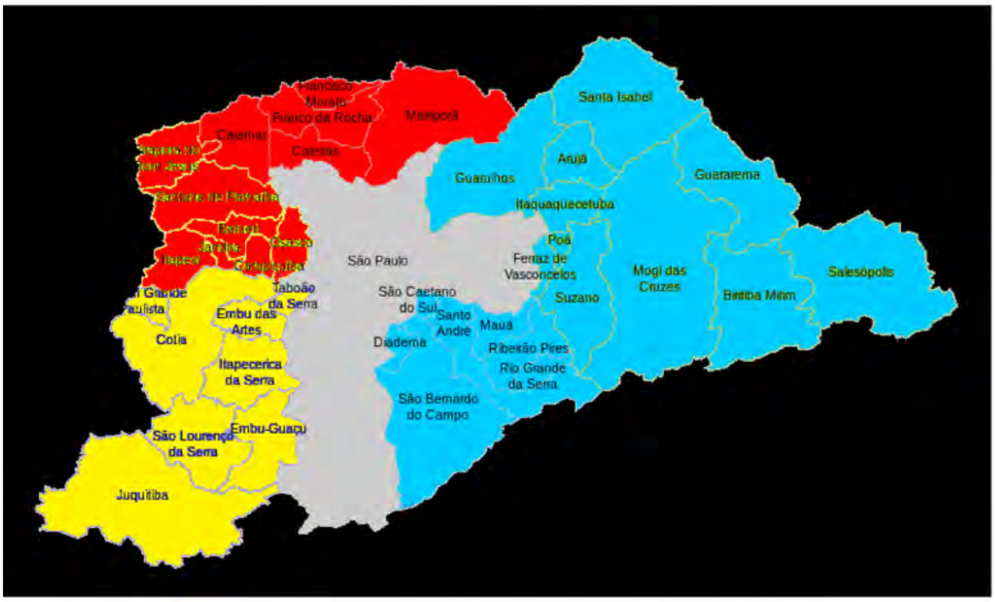

Figure 6 - Division of the outer São Paulo area into zones

Figure 7 - Campinas Metropolitan Area and São Paulo State

Table 32 summarizes this division.

Table 32 - Summary of geographic divisions

Zone	Zone ID	Zone components
São Paulo - Central	1	Central districts of the São Paulo Municipality
São Paulo – East 1	2	Inner eastern districts of the São Paulo Municipality
São Paulo – East 2	3	Outer eastern districts of the São Paulo Municipality
São Paulo – North 1	4	Northeastern districts of the São Paulo Municipality
São Paulo – North 2	5	Northwestern districts of the São Paulo Municipality
São Paulo – West	6	Western districts of the São Paulo Municipality
São Paulo – South 1	7	Inner southern districts of the São Paulo Municipality
São Paulo – South 2	8	Outer southern districts of the São Paulo Municipality
Campinas Metropolitan Area	9	Municipalities of Americana, Artur Nogueira, Campinas, Cosmópolis, Engenheiro Coelho, Holambra, Hortolândia, Indaiatuba, Itatiba, Jaguariúna, Monte Mor, Nova Odessa, Paulínia, Pedreira, Santa Bárbara do Oeste, Santo Antônio de Posse, Sumaré, Valinhos and Vinhedo
São Paulo Metro North	10	Municipalities of Barueri, Caieiras, Cajamar, Carapicuíba, Francisco Morato, Franco da Rocha, Itapevi, Jandira, Mairiporã, Pirapora do Bom Jesus, Osasco and Santana de Parnaíba
São Paulo Metro Southwest	11	Municipalities of Cotia, Embu das Artes, Embu-Guaçu, Itapecerica da Serra, Juquitiba, São Lourenço da Serra, Taboão da Serra and Vargem Grande Paulista
São Paulo Metro Southeast	12	Municipalities of Arujá, Biritiba Mirim, Diadema, Ferraz de Vasconcelos, Guarulhos, Gurararema, Itaquaquecetuba, Mauá, Mogi das Cruzes, Poá, Ribeirão Pires, Rio Grande da Serra, Salesópolis, Santa Isabel, Santo André, São Bernardo do Campo, São Caetano do Sul and Suzano,
São Paulo State	13	Municipalities in São Paulo state not part of the Campinas and São Paulo metropolitan areas

After the passengers were segmented into zones it became possible to calculate the average time spent by passengers from each zone to each airport, as displayed by Table 33. The averages were based on the observations shown on Table 34.

Table 33 - Access times by zone

	Access Time (minutes)			
Zone	**GRU**	**CGH**	**VCP**	**Average**
São Paulo - Central	59.0	28.2	140.0	**41.8**
São Paulo – East 1	50.5	39.3	150.0	**47.5**
São Paulo – East 2	38.1	54.5	180.0	**43.1**
São Paulo – North 1	42.1	39.8	120.0	**41.7**
São Paulo – North 2	52.8	36.4	90.0	**45.2**
São Paulo – West	61.8	26.3	97.3	**41.8**
São Paulo – South 1	61.2	16.1	93.3	**33.8**
São Paulo – South 2	63.7	32.4	62.0	**46.3**
Campinas Metropolitan Area	115.6	121.0	28.7	**75.6**
São Paulo Metro North	82.2	53.4	84.4	**70.3**
São Paulo Metro Southwest	81.2	50.0	120.0	**68.4**
São Paulo Metro Southeast	53.6	50.7	60.0	**52.7**
São Paulo State	155.4	112.0	110.4	**141.6**
Total	**80.9**	**35.6**	**67.2**	**60.1**

Table 34 - Number of observations by zone and airport for access time calculations

	Observations			
Zone	**GRU**	**CGH**	**VCP**	**Average**
São Paulo - Central	463	598	2	**1063**
São Paulo – East 1	231	100	1	**332**
São Paulo – East 2	66	29	-	**95**
São Paulo – North 1	194	136	2	**332**
São Paulo – North 2	105	96	1	**202**
São Paulo – West	934	1290	22	**2246**
São Paulo – South 1	431	682	3	**1116**
São Paulo – South 2	366	464	5	**835**
Campinas Metropolitan Area	214	40	221	**475**
São Paulo Metro North	211	156	9	**376**
São Paulo Metro Southwest	13	9	-	**22**
São Paulo Metro Southeast	500	251	1	**752**
São Paulo State	929	256	169	**1354**
Total	**4657**	**4107**	**436**	**9200**

An analysis of Table 34 shows that for some, especially for Viracopos, the number of observations is too small (under 30), and in two cases, there were no observations. In the cases in which no observation was available access time was estimated by calculating road access time based on distance and speed, and then multiplying by an 1.5 factor to compensate for congestion. As for cases in which the number of observations was small, given that they did not substantially deviate from the above method, they were kept, although this is likely to have a negative impact on the demand division model.

A third key attribute of passengers is experience with each airport. The questionnaire applied by FIPE (2009) includes a question on the number of times an air route was taken by a passenger in the last 12 months. This can be considered an indicator of the passenger's experience with the airport. There is no ready information, however, regarding experience with other airports.

Again it becomes necessary to estimate these parameters from the database. In order to do this a three-step approach was adopted. First, the average experience of passengers at each zone was calculate, yielding the results on Table 35.

Table 35 - Average passenger experience by zone

	Number of times route was taken in 12 months			
Zone	**GRU**	**CGH**	**VCP**	**Average**
São Paulo - Central	3.66	9.99	2.50	**7.21**
São Paulo – East 1	3.01	8.86	0.00	**4.77**
São Paulo – East 2	1.17	5.97		**2.63**
São Paulo – North 1	3.89	9.51	0.00	**6.18**
São Paulo – North 2	2.27	12.85	0.00	**7.29**
São Paulo – West	4.04	12.01	2.73	**8.60**
São Paulo – South 1	3.08	12.17	1.00	**8.63**
São Paulo – South 2	4.03	10.71	1.80	**7.72**
Campinas Metropolitan Area	1.46	8.40	7.69	**4.92**
São Paulo Metro North	2.84	9.99	8.63	**5.93**
São Paulo Metro Southwest	1.15	16.63		**7.05**
São Paulo Metro Southeast	4.29	10.11	20.00	**6.24**
São Paulo State	2.80	8.33	4.62	**4.07**
Total	**3.37**	**10.96**	**6.07**	**6.89**

Based on this data the next step was calculating for each passenger an experience coefficient. The experience coefficient denotes how much the passenger is experienced with an airport compared to peers in his region, and is given by:

$$experience\ coefficient = \frac{passenger\ experience}{average\ zone\ experience\ with\ airport} \qquad (16)$$

The final step assumes that, the more experienced a passenger is with is airport of choice, the more he is also likely to have some experience with the other airports. Based on this the final operation is multiplying the experience coefficient by the average experience of zone peers with the non-chosen airports, thus yielding an estimate of passenger experience for the airports not chosen.

6.6 Description of models tested

6.6.1 Modeling Approach definition

The literature review performed indicates that access depends chiefly on accessibility, flight offer and passenger experience variables, such as:

- Accessibility: Door-to-door airport access time, distance
- Flight offer: Frequency of direct flights, frequency of indirect flights (flights with connections)
- Passenger experience: Number of times a passenger used a given airport

Literature review also made it clear that the preferred approaches for modeling were multinomial logit and nested logit.

In the case of this study, following the approach adopted by Moreno and Müller (2002, 2003), the chosen approach was to model choice using a multinomial logit model, using frequency, experience and access time as prediction variables.

As explained in item 6.5, the database, after adjustment, includes access times for each airport, allowing the use of an access time variable (AT), and direct weekly frequency (DWF) data for each airport, for each passenger. Experience, as the yearly number of times a route is taken, is available as an actual number of the chosen airports and as an estimation for the non-chosen airports.

The model variables therefore are:

- Access Time (given in minutes)
- Weekly Direct Frequencies
- Number of times route was taken in the last 12 months

Another aspect to be considered in model definition is segmentation, a problem for which there is an important trade-off that must be taken into account.

On one hand, different segments have different choice behaviors. Literature review suggests two divisions.

The first is a division between business and non-business passengers. Table 36 shows how these two groups of passengers chose airports:

Table 36 - Passenger choice segment by travel purpose

	Airport Choice			
Zone	**GRU**	**CGH**	**VCP**	**Total**
Non-business	2952	1234	198	**4024**
Business	2065	2873	238	**5176**
Total	**4657**	**4107**	**436**	**9200**

It is possible to see, on Table 36, that business passengers tend to prefer the Congonhas as opposed to Guarulhos Airport. This is understandable, as Congonhas is the airport closest to the business center. This segmentation seems, however, to have a limited effect on the Viracopos airport.

The second is a division along regional or place of residence lines. Table 37 shows the split of passenger choice by region:

Table 37 - Passenger choice segmented by geographic zones

	Airport Choice			
Zone	**GRU**	**CGH**	**VCP**	**Total**
São Paulo - Central	463	598	2	**1063**
São Paulo – East 1	231	100	1	**332**
São Paulo – East 2	66	29	-	**95**
São Paulo – North 1	194	136	2	**332**
São Paulo – North 2	105	96	1	**202**
São Paulo – West	934	1290	22	**2246**
São Paulo – South 1	431	682	3	**1116**
São Paulo – South 2	366	464	5	**835**
Campinas Metropolitan Area	214	40	221	**475**
São Paulo Metro North	211	156	9	**376**
São Paulo Metro Southwest	13	9	-	**22**
São Paulo Metro Southeast	500	251	1	**752**
São Paulo State	929	256	169	**1354**
Total	**4657**	**4107**	**436**	**9200**

An analysis of Table 37 shows, as expected, that zones closest to the airports have greater attractiveness, a fact that is likely to be reflected by the access time variable. This is not the case, however, for the non-metropolitan São Paulo State. Even though access time varies little between airports for these cities there is a markedly above average preference for the Viracopos airport. This suggests that this area has a different response to access time than other areas. This suggests considering two regional segments – a metropolitan segment and a non-metropolitan segment.

This is very relevant for this study as Viracopos and the impact of access time are its main concerns. Table 38 shows the choice split between the two regional segments.

Table 38 - Choice split between passenger originating from São Paulo and Campinas metropolitan areas and from other São Paulo state municipalities

	Airport Choice			
Zone	**GRU**	**CGH**	**VCP**	**Total**
Metropolitan	3728	3851	267	**7846**
Non-Metropolitan	929	256	169	**1354**
Total	**4657**	**4107**	**436**	**9200**

The decision of splitting into segments, however, depends on the second element of the trade-off: the number of observations. A low number of observations compromises the effectiveness of model. Moreno and Müller (2002), commenting on the number of observations normally used by literature, indicate that a number of observations around 2000 is satisfactory for a three-airport choice problem. In this study, in spite of the good number of observations, very few passengers chose the Viracopos airport, only 436 out of 9200. Moreover, as discussed in the section concerning data, several regions are very underrepresented in this universe, and excessive segmentation might mean that their behavior is altogether lost.

Considering this trade-off only one segmentation was used – the regional segmentation. This decision is justified by the greater interest that changes in access time to Viracopos, among all factors, offer to this study.

The models to be fitted, therefore, are initially two:

- A multinomial logit model with frequency, experience and access time variables for the São Paulo and Campinas metropolitan areas
- A multinomial logit model with frequency, experience and access time variables for municipalities in the São Paulo state outside the São Paulo and Campinas metropolitan areas

6.7 Fitting of intercept model for comparison purposes

With models defined, the next step is to fit the models to the data. The multinomial conditional logit model can be fitted by Stata using the command asclogit.

As discussed in item 6.1, one of the ways to verify the usefulness of a conditional logit is to calculate McFadden's R-squared. This metric depends both on the model tested and on a version of the model with no explanatory variables – the intercept model. In the intercept model the intercept coefficient attempts to approximate the effect of all variables that could influence choice, but were not specified.

In this section, therefore, intercept models will be estimated for the two regional segments.

6.7.1 Model for the São and Campinas Metropolitan Areas

Table 39 shows the results of fitting the conditional logit model to the first regional dataset using asclogit in Stata, using Congonhas as base choice.

Table 39 - Coefficients and standard deviations of intercept model for the São Paulo and Campinas metropolitan areas with Congonhas as base choice

Airport Choice	Coefficients	Value (Standard Error)		P>|z|
GRU	------------- Base Alternative -------------			
CGH	Constant	0.032461 (0.2029764)		0.158
VCP	Constant	-2.636380 (0.0633527)	***	0.000
Log-likelihood		**-6417.3652**		

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Table 39 is subject to the following interpretations:

All results the coefficients given are the confidents applicable to GRU and VCP when their choice probability is compared to the choice probability for CGH. Applying the data to the set of equations 8-10 for the conditional logit model this means that:

$$\ln \frac{p_{GRU}}{p_{CGH}} = \beta_{0,GRU} = 0.032461 \quad (18)$$

$$\ln \frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} = -2.636380 \quad (19)$$

This means that, when the most likely values for the intercept coefficients are considered, passengers are slightly less likely to travel from CGH than GRU, and far less likely to travel from VCP than GRU. There is one important caveat, however: as the p>|z| test shows, there is low confidence in affirming that the intercept for Congonhas is not null. In practice this means that it is not possible to differentiate probabilities of choice between CGH and GRU without seeking for explanatory variables.

The final important number shown by Table 39 is the logarithm of the likelihood function given by equation 15, which equals -6417.3652. This is the number that will be used to compute McFadden's R^2 for subsequent models for the São Paulo and Campinas metropolitan areas.

6.7.2 Model for the São Paulo State outside the São Paulo and Campinas Metropolitan Areas

Table 40 shows the results of fitting the conditional logit model to the second regional dataset using asclogit in Stata, using GRU as base choice.

Table 40 - Coefficients and standard deviations of intercept model outside the São Paulo and Campinas metropolitan areas

Airport Choice	Coefficients	Value (Standard Error)		P>|z|
GRU		------------- Base Alternative -------------		
CGH	Constant	-1.288931 (0.0705881)	***	0.000
VCP	Constant	-1.704210 (0.0836277)	***	0.000
Log-likelihood		**-1128.0436**		

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Table 40 is subject to the following interpretations:

All results the coefficients given are the confidents applicable to GRU and VCP when their choice probability is compared to the choice probability for CGH. Applying the data to the set of equations 8-10 for the conditional logit model this means that:

$$\ln\frac{p_{CGH}}{p_{GRU}} = \beta_{0,CGH} = -1.288931 \tag{20}$$

$$\ln\frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} = -1.704210 \tag{21}$$

This means that, when the most likely values for the intercept coefficients are considered, passengers less likely to travel from CGH than GRU, and also less likely to travel from VCP than GRU, though in a lesser degree than in the previous model. In this case, as the p>|z| test shows, there is very high confidence in the intercepts not being null, which means that the estimated probabilities of choice are at least directionally correct.

The logarithm of the likelihood function equals -1288.0429. This is the number that will be used to compute McFadden's R^2 for subsequent models for the cities in the São Paulo state outside the São Paulo and Campinas metropolitan areas.

6.8 Models based on frequency, experience and access

With intercept models estimated as comparison the next step is to estimate models with the explanatory variables proposed.

6.8.1 Model for the São Paulo and Campinas Metropolitan Areas

Table 41 shows the results of fitting the conditional logit model with experience, access time and frequency variables to the first regional dataset using asclogit in Stata, using GRU as base choice.

Table 41 - Coefficients and standard deviations of frequency experience and access time model for the São Paulo and Campinas metropolitan areas

Airport Choice	Coefficients	Value (Standard Error)		P>|z|
GRU	------------- Base Alternative -------------			
CGH	Constant	-0.1367772 (0.0371713)	***	0.000
VCP	Constant	-1.603345 (0.071713)	***	0.000
All Choices	Experience	0.0005554 (0.0017613)		0.753
	Access Time	-0.0187811 (0.0007467)	***	0.000
	Direct Frequencies	0.0123996 (0.0004629)	***	0.000
Log-likelihood	**-4503.0339**			

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Data from Table 41 is subject to the following interpretations:

As was the case with the intercept model the coefficients given are the confidents applicable to GRU and VCP when their choice probability is compared to the choice probability for CGH. Each choice has its own intercept.

The coefficients for the experience, access time and frequency, however work in a different manner. The coefficients are global, but each choice has its own variable for each of the three factors.

Applying the data to the set of equations 8-10 yields:

$$\ln\frac{p_{CGH}}{p_{GRU}} = \beta_{0,CGH} + \beta_{exp} \times exp + \beta_{atime} \times atime + \beta_{dwf} \times dwf = -0.1367772 + 0.0005554 \times exp - 0.0187811 \times atime + 0.0123996 \times dwf \tag{22}$$

$$\ln\frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} + \beta_{exp} \times exp + \beta_{atime} \times atime + \beta_{dwf} \times dwf = -1.603345 + 0.0005554 \times exp - 0.0187811 \times atime + 0.0123996 \times dwf \tag{23}$$

The logarithm of the likelihood function equals -4503.0339. This, coupled with the likelihood from the intercept model, makes it possible to compute McFadden's R^2:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} = 1 - \frac{4503.0039}{6417.3652} = 0.2983 \tag{24}$$

The 0.2983 value for McFadden's R^2 suggests a fair, though not excellent, fit.

As expected, longer access times have a significant negative factor on choice, while higher frequencies encourage airport use. The p>|z| test shows that the coefficients for these factors are not null, as well as the intercepts.

The experience variable, however, was not significant at the 95% or even at the 90% level, meaning that it offers no reliability for estimation. This makes it necessary to fit a model without the experience variable.

The fact that the experience variable was not significant is of note, as it runs counter to the conclusion achieved by Moreno and Müller (2002, 2003).

6.8.2 Model for the São Paulo State outside the São Paulo and Campinas Metropolitan Areas

Table 42 shows the results of fitting the conditional logit model with experience, access time and frequency variables to the first second dataset using asclogit in Stata, using GRU as base choice.

Table 42 - Coefficients and standard deviations of frequency experience and access time model outside the São Paulo and Campinas metropolitan areas

Airport Choice	**Coefficients**	**Value (Standard Error)**		**P>|z|**
GRU	------------- Base Alternative -------------			
CGH	Constant	-0.7871112 (0.0963657)	***	0.000
VCP	Constant	-1.179225 (0.1082505)	***	0.000
All Choices	Experience	0.0085032 (0.0065405)		0.194
	Access Time	-0.0004319 (0.0006426)	***	0.502
	Direct Frequencies	0.0050551 (0.0013076)	***	0.000
Log-likelihood		**-1010.3285**		

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Data from Table 42 is subject to the following interpretations:

The coefficients work in the same manner as in the previous model – constants for each choice and global coefficients for choice-specific variables

Applying the data to the set of equations 7-9 yields:

$$\ln\frac{p_{CGH}}{p_{GRU}} = \beta_{0,CGH} + \beta_{exp} \times exp + \beta_{atime} \times atime + \beta_{dwf} \times dwf = -0.7871112 - 0.0085032 \times exp - 0.0004319 \times atime + 0.0050551 \times dwf \quad (24)$$

$$\ln\frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} + \beta_{exp} \times exp + \beta_{atime} \times atime + \beta_{dwf} \times dwf = -1.179225 - 0.0085032 \times exp - 0.0004319 \times atime + 0.0050551 \times dwf \quad (25)$$

The logarithm of the likelihood function equals -1010.3285. This, coupled with the likelihood from the intercept model, makes it possible to compute McFadden's R^2:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} = 1 - \frac{1010.3285}{1128.0429} = 0.1044 \qquad (26)$$

The 0.1044 value for McFadden's R^2 is a low value. This indicates that important factor not included in the model could improve it.

Higher frequencies encourage airport use, and are significant to under the 1% level, as the p>|z| test shows. The coefficients are significant as well.

The experience and frequency variables, however, were not significant. This makes it necessary to fit a model without the experience and frequency variables.

The fact that the experience variable was not significant again is contrary to the conclusion achieved by Moreno and Müller (2002, 2003).

6.9 Models based on frequency and access

6.9.1 Model for the São Paulo and Campinas Metropolitan Areas

Table 43 shows the results of fitting the conditional logit model with access time and frequency variables to the first dataset using asclogit in Stata, using GRU as base choice.

Table 43 - Coefficients and standard deviations of frequency model for the São Paulo and Campinas metropolitan areas

Airport Choice	Coefficients	Value (Standard Error)		P>|z|
GRU	-------------- Base Alternative --------------			
CGH	Constant	0.1383462 (0.0354405)	***	0.000
VCP	Constant	-1.588958 (0.071186)	***	0.000
All Choices	Access Time	-0.0189314 (0.0007398)	***	0.000
	Direct Frequencies	0.0123278 (0.0004594)	***	0.000
Log-likelihood		**-4535.9802**		

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Data from Table 43 is subject to the following interpretations:

As was the case with models tested previously the coefficients are of two kinds: choice-specific intercepts for each choice and global coefficients for choice-specific access time and frequency variables.

Applying the data to the set of equations 8-10 yields:

$$\ln\frac{p_{CGH}}{p_{GRU}} = \beta_{0,CGH} + \beta_{atime} \times atime + \beta_{dwf} \times dwf = 0.1383462 - 0.0189314 \times atime + 0.0123278 \times dwf \quad (27)$$

$$\ln\frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} + \beta_{atime} \times atime + \beta_{dwf} \times dwf = -1.588958 - 0.0189314 \times atime + 0.0123278 \times dwf \quad (28)$$

The logarithm of the likelihood function equals -4535.9802. This, coupled with the likelihood from the intercept model, makes it possible to compute McFadden's R^2:

$$R^2 = 1 - \frac{\ln\hat{L}(M_{Full})}{\ln\hat{L}(M_{Intercept})} = 1 - \frac{4535.9802}{6417.3652} = 0.2932 \quad (29)$$

The 0.2932 value for McFadden's R^2 suggests that the frequency and access time model has a fair fit with data.

As expected, longer access times have a significant negative factor on choice, while higher frequencies encourage airport use. The p>|z| test shows that the coefficients for these factors are not null, as well as the intercepts. This indicates that this is a feasible model for airport choice in the São Paulo and Campinas Metropolitan Areas.

6.10 Models based on frequency

6.10.1 Model for the São Paulo State outside the São Paulo and Campinas Metropolitan Areas

Table 44 shows the results of fitting the conditional logit model with frequency variables to the second regional dataset using asclogit in Stata, using GRU as base choice.

Table 44 - Coefficients and standard deviations of frequency model outside the São Paulo and Campinas metropolitan areas

Airport Choice	Coefficients	Value (Standard Error)		P>|z|
GRU	------------- Base Alternative -------------			
CGH	Constant	-0.8266825 (0.0012992)	***	0.000
VCP	Constant	-1.183999 (0.103355)	***	0.000
All Choices	Direct Frequencies	0.0049873 (0.0012992)	***	0.000
Log-likelihood		**-1017.7736**		

Note: * = significantly different from zero at 10% level; ** = significantly different from zero at 5% level; *** = significantly different from zero at 1% level

Data from Table 44 is subject to the following interpretations:

As was the case with models tested previously the coefficients are of two kinds: choice-specific intercepts for each choice and global coefficients for choice-specific frequency variables.

Applying the data to the set of equations 8-10 yields:

$$\ln\frac{p_{CGH}}{p_{GRU}} = \beta_{0,CGH} + +\beta_{dwf} \times dwf = -0.8266825 + 0.0049873 \times dwf \quad (31)$$

$$\ln\frac{p_{VCP}}{p_{CGH}} = \beta_{0,VCP} + \beta_{dwf} \times dwf = -1.183999 + 0.0049873 \times dwf \quad (32)$$

The logarithm of the likelihood function equals -1017.7736. This, coupled with the likelihood from the intercept model, makes it possible to compute McFadden's R^2:

$$R^2 = 1 - \frac{\ln \hat{L}(M_{Full})}{\ln \hat{L}(M_{Intercept})} = 1 - \frac{1017.7736}{1128.0229} = 0.0978 \quad (33)$$

The 0.0978 value for McFadden's R^2 suggests that the frequency and access time model has a limited descriptive power.

As expected, higher frequencies encourage airport use. The $p>|z|$ test shows that the coefficient for frequency is not null, as well as the intercepts.

Even though the model has a poor fit, it offers the important insights that 1. Access times are of little relevance for passengers outside São Paulo and Campinas; 2. Frequency is still an important factor for choice. This will be important for the final demand estimation.

7 Demand estimation for the São Paulo area airports considering changes in access time

With models built both for the aggregate demand for the São Paulo area and for airport choice within the region, it is possible to verify the impacts of access time on demand.

This section will discuss two scenarios: one for 2030 considering unchanged access times, considered as a baseline, and another considering reduced access times to Viracopos as consequence of a new, faster modal, such as the TAV.

Before discussing the scenarios themselves, however, it is necessary to define common assumptions regarding the relationship between departures from the São Paulo state from Viracopos, Congonhas and Guarulhos, which were modeled by this study, and overall passenger traffic at the airports.

7.1.1 Assumptions for passenger traffic in São Paulo in 2030

The first assumption made is that departures from the São Paulo state can be considered equal to departures during a 12 month period.

A second assumption is assuming that any passengers originating outside the São Paulo state fly into the airport and therefore can be considered passengers in connection. This is supported by the fact that approximately 2% of passengers at the São Paulo who were not connection passengers originated outside the São Paulo State.

A third assumption has to made regarding the demand split between passengers with the São Paulo and Campinas metropolitan areas. A precise study would require analysis of the economic and demographic trends for each region, a task outside the scope of this study. Considering however, the forecast in demand growth for the São Paulo state (McKinsey 2009) and a benchmark share of connection flights this study will use the shares shown on Table 45.

Table 45 shows the split between connections, flights originating from the São Paulo and Campinas Metropolitan Areas and flights from other parts of the São Paulo state. Passenger totals use the aggregate demand baseline given by the adjusted McKinsey (2009) estimates.

Table 45 - Assumptions for demand sources at the São Paulo Area airports for 2030

	Connection	São Paulo + Campinas Areas	Other São Paulo State	Total
Split	33,3%	50%	16.7%	**100.0%**
Neutral baseline (millions of passengers)	56.0	84.0	28.0	**168.1**
Conservative baseline (millions of passengers)	42.9	64.4	21.4	**128.7**

Based on these assumptions, the approach for estimating demand was:

1. Defining assumptions for frequency and accessibility for CGH, GRU and VCP in 2030
2. Based on the final demand choice models, predict probabilities of choice for a forecasted “average profile” passenger
3. Multiply probabilities of choice by expected demand for segment, taking into account assumptions about demand split in from Table 45

Estimates will be done considering the neutral and conservative aggregate demand estimates.

7.1.2 Reference scenario – demand without changes in access time

The first scenario projected attempts to estimate demand without the TAV. Its main assumptions are:

- Frequencies at Congonhas are constant from 2008, due to capacity constraints
- The sum of frequencies for the three airports grows in tandem with the aggregate demand forecast
- Both domestic and international frequencies at GRU and CGH are proportional to maximum capacity (50 million and 80 million pax/year, respectively)
- International and Domestic split is 10%
- Access times are constant compared to 2008 for all airports

In this scenario the forecast demand, in 2030, for each airport is that shown in Table 46, for the neutral baseline, and in Table 47 for the conservative baseline.

Table 46 - Demand for São Paulo Area airports in 2030 in a constant access time scenario, considering neutral baseline

Zone	Zone	VCP	GRU	CGH	Total
São Paulo and Campinas Metro (Dept+Arr)	Split	53.6%	42.8%	3.6%	**100%**
	Passengers (millions)	45.0	35.9	3.1	**84.0**
Other São Paulo state (Dept+Arr)	Split	35.1%	55.8%	9.1%	**100%**
	Passengers (millions)	9.8	15.6	2.5	**28.0**
Connections	Split	48.9%	46.1%	5.0%	**100%**
	Passengers (millions)	27.7	25.8	2.7	**56.0**
Total	**Split**	**48.9%**	**46.1%**	**5.0%**	**100.0%**
	Passengers (millions)	**82.3**	**77.4**	**8.4**	**168.1**

Table 47 - Demand for São Paulo Area airports in 2030 in a constant access time scenario, considering the conservative baseline

Zone	Zone	VCP	GRU	CGH	Total
São Paulo and Campinas Metro (Dept+Arr)	Split	53.6%	42.8%	3.6%	**100%**
	Passengers (millions)	34.5	27.5	2.4	**64.4**
Other São Paulo state (Dept+Arr)	Split	35.1%	55.8%	9.1%	**100%**
	Passengers (millions)	7.4	12.0	2.0	**21.4**
Connections	Split	48.9%	46.1%	5.0%	**100%**
	Passengers (millions)	21.0	19.8	2.1	**56.0**
Total	**Split**	**48.9%**	**46.1%**	**5.0%**	**100.0%**
	Passengers (millions)	**63.0**	**59.4**	**6.4**	**128.8**

7.1.3 Demand with TAV to Viracopos

For the scenario with TAV the assumptions adopted were the same as those for the reference scenario, with a change in the access time to Viracopos, which is reduced to 55 minutes. This justified by an estimated time of 70 minutes for TAV access, split between 48 between origin and the TAV station at Campo de Marte and 22 min between the TAV station and the airport (Halcrow, 2009).

In this scenario the forecasted demand, in 2030, for each airport is that shown in Table 48 for the neutral baseline and in Table 49 for the conservative baseline:

Table 48 - Demand for São Paulo Area airports in 2030 in a scenario with TAV, considering the neutral baseline

Zone	Zone	VCP	GRU	CGH	Total
São Paulo and Campinas Metro (Dept+Arr)	Split	70.8%	26.9%	2.3%	**100%**
	Passengers (millions)	59.5	22.6	1.9	**72.7**
Other São Paulo state (Dept+Arr)	Split	35.1%	55.8%	9.1%	**100%**
	Passengers (millions)	9.8	15.6	2.5	**28.0**
Connections	Split	61.9%	34.2%	4.0%	**100%**
	Passengers (millions)	34.7	19.1	2.2	**48.4**
Total	**Split**	**61.9%**	**34,2%**	**4.0%**	**100.0%**
	Passengers (millions)	**104.0**	**57.4**	**6.7**	**168.1**

Table 49 - Demand for São Paulo Area airports in 2030 in a scenario with TAV, considering the conservative baseline

Zone	Zone	VCP	GRU	CGH	Total
São Paulo and Campinas Metro (Dept+Arr)	Split	70.8%	26.9%	2.3%	**100%**
	Passengers (millions)	45.6	17.3	1.5	**64.4**
Other São Paulo state (Dept+Arr)	Split	35.1%	55.8%	9.1%	**100%**
	Passengers (millions)	7.4	12.0	2.0	**21.4**
Connections	Split	61.9%	34.2%	4.0%	**100%**
	Passengers (millions)	26.6	14.7	1.7	**42.9**
Total	**Split**	**61.9%**	**34,2%**	**4.0%**	**100.0%**
	Passengers (millions)	**79.6**	**44.0**	**5.1**	**128.8**

The scenario with TAV forecasts an increase between 16.6 and 21.7 million passengers due to the TAV. Without the TAV, in the conservative scenario Guarulhos is nearly at full utilization (60 million passengers, as shown in Table 2) and far above it in the neutral scenario. Viracopos, on the other hand, has some expansion potential left even in the more aggressive scenario.

In the scenarios with TAV using both the neutral and conservative baselines Guarulhos has demand bellow maximum capacity. In the conservative demand forecast Viracopos is under full capacity, but over in the neutral demand forecasts.

In all scenarios Congonhas is below capacity. This suggests that the assumption of constant frequencies may have been too restrictive, and that the airport may possibly relieve excess some demand from Viracopos and Guarulhos. This does not change, however, the fact that in the more aggressive demand forecast the aggregate demand exceeds the combined maximum capacities of the airports, implying that a new airport is needed.

8 Policy and research recommendations

8.1 Policy recommendations

In the definition of the demand baseline for the São Paulo area it was pointed out that the higher estimate of 168 million passengers per year estimate of aggregate demand implied more than full utilization of the airports, even considering capacity expansion to maximum levels. This suggests that, if demand follows in the next years a path similar to the optimistic scenario, a new airport will be necessary.

Correia, Nyama and Nogueira (2011) suggested a set of locations at or near the São Paulo municipality as places for a new airport.

Also, as shown in the demand scenarios for each airport in 2030, the TAV is helpful to ensure that Viracopos reaches its full potential in time to avoid critical investment resistant overcrowding in Guarulhos. The lack of a fast access to Viracopos results in greater pressure to expand Guarulhos, an airport where due to constraints each increment in capacity tends to be more expensive. The combination of higher use of Viracopos and less pressure on Guarulhos can yield economic benefits.

It is important to take into account, however, that the total cost of the TAV between Rio de Janeiro, São Paulo and Campinas is over USD 30 Billion (Halcrow, 2009), which might not be justified only based on the benefits of better access to Viracopos. This suggests the possibility of considering the access to Viracopos as a separate project.

In the estimates of reduced access time the main component of time was access to the train station, over twice as long as the train trip itself. This indicates that a more economical solution might be optimizing station location while investing on a cheaper, lower velocity train technology, yielding the same returns with lower investment

Summarizing, the policy recommendations are:

- Building, if demand growth keeps up with projections, a new airport at or near São Paulo
- Considering that an alternative, faster access mode to Viracopos is a priority, this access mode not necessarily being the TAV

- Investigate the possibility of building a cheaper, slower but better located train system connecting São Paulo and the Viracopos airport
- Consider judging the economic merits of the TAV based on sections, as the Viracopos-São Paulo section has implications on airport investments

8.2 Research recommendations

The first part of this study sought to estimate aggregate demand for the São Paulo area airports through adjustments to the model developed by McKinsey (2009). The adjustments demonstrate the importance in such estimations of:

- Taking into account inequality reduction effects, as Brazilian income inequality is falling
- Paying attention to changes in the behavior of demand over time – this might indicate that variables or their coefficients changed over time
- Keep in mind that yield is a volatile indicator, and might fall abruptly due to competition and the entrance of low cost companies, such as Azul

In the second part of the study a discrete choice model was used to split demand between the São Paulo airports. This analysis was compromised by a limited number of questionnaires at Viracopos, the lack of a more refined access time model, the lack of information on indirect frequencies at airports and the difficulty in evaluating the experience of passengers with non-chosen airports.

These difficulties, together with the continuing possibility of a capacity bottleneck with continued demand growth at the São Paulo area airports make it necessary to achieve a better understanding of airport choice in the region. Studies in this area should follow several recommendations:

- There is a need for an extensive survey of the passengers in the area, with a concern collecting a sufficient number of observations at Viracopos
- These questionnaires, having in mind the choice possibilities, should evaluate the passenger’s experience with the three airports
- Future studies should incorporate transportation models in order to estimate with precision access times through different transportation modes

Moreover, the access benefits of the TAV represent an important economic tradeoff. Future studies might investigate this tradeoff in order to evaluate the economic benefit both of the TAV project and the TAV section between Viracopos and São Paulo.

9 Conclusions

This work attempted to understand the effects of changes in the access time to Viracopos, especially regarding the TAV. In order to accomplish this, two models were developed and then interacted: a logarithmic regression model for demand of the Guarulhos, Viracopos and Congonhas airports taken as a unit and a choice model between these three airports using a conditional logistic regression.

The linear regression model set out to improve on the model developed by McKinsey (2009). It was found out that adjustments were capable of significantly improving the model and substantially reducing forecast errors occurred in the first years after the forecast. Changes in GDP and yield values proved important. Even more significant was the effect of income inequality, incorporated into the model by means of a metric developed by Sen (1982). This marked effect suggests that income inequality should be taken into account while evaluating airport demand.

The results yielded by this reviewed model were used as a reference by this study.

The next step was fitting choice models to the passenger survey done by FIPE (2009) in order to gain forecasting ability to passenger choice. Two models were made for different geographic zones. Weekly direct frequencies and access time were significant to the 95%, but not experience. Experience not being a significant variable contrasted with the findings of Moreno and Müller (2003).

With the demand and choice models set up a demand forecast was made for 2030 for each of the three São Paulo area airports. It was found out that the TAV could have a considerable influence on the demand of Viracopos, and that it played an important role in optimizing the use of airport capacity in the region. It was also observed that there existed a possibility of all airports exceeding capacity, demanding that a new airport be built before 2030.

The work recommended that a fast access be built to Viracopos, due to its influence on airport choice. This access does not need to be the TAV – a medium speed train with well-localized terminals might be more cost effective. The work also recommended that the merits of the TAV for the São Paulo-Viracopos section be judged separately from the rest of the project.

The work concluded with suggestions for research. It is important to follow-up on this work, considering the need of developing a new, more extensive dataset and a more precise control of access times indicators. It is also important to better evaluate the economic trade-offs of the TAV considering the gains on airport accessibility shown by this study.

References

ANAC – Agência Nacional de Aviação Civil. **Relatório de Tarifas Aéreas, 25th edition**. Brasília:ANAC. 2012.

Ashford N.; Becherman M. **Passengers' Choice of Airport: An application of the multinomial Logit Model**. Transportation Research Record 1147, pp.1-5, 1987.

Augustinus, J. G.; Demakopoulos, S.A. **Air Passenger Distribution Model for a Multiterminal Airport System**, Transportation Research Record 673, pp. 176-180, 1978.

Bierens, Herman J. **The logit model: estimation, Testing and Interpretation.** Pennsyvania State University, Department of Economics, 2008. Available at: <http://econ. la.psu.edu/~ hbierens/ML_LOGIT. PDF> Access in nov. 7th, 2012.

Condé, M. **Estudo e previsão de demanda aeroportuária para a cidade do Rio de Janeiro.** Journal of Transport Literature, vol. 5, n. 1, pp. 161-183, 2011.

Correia, A. R.; Niyama, L. E.; Nogueira, S. A. F. **Estimativa da distribuição da demanda na região metropolitana de São Paulo com cenários de um novo aeroporto.** Journal of Transport Literature, vol. 5, n. 2, pp. 44-62, 2011.

Cushing, C.; Cushing, B. **Conditional Logit, IIA, and Alternatives for Estimating Models of Interstate Migration**. Paper presented at the 46th annual meeting of the Southern Regional Science Association. Charleston, SC, 2007. Available at <http://rri.wvu.edu/pdffiles/wpcushing2007-4.pdf> Access in nov. 8th, 2012.

FIPE – Fundação Instituto de Pesquisas Econômicas – **Caracterização e dimensionamento da matriz origem destino do transporte aéreo no Brasil – Plano de Pesquisa**. 2009, Available at <http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/empresa/pesquisa/chamada3/20100125-Plano_de_Pesquisa.pdf > Access in nov. 11th, 2012.

Gastwirth, J. L. **A General Definition of the Lorenz Curve**. Econometrica, pp. 1037-1039, 1971.

________. The estimation of the Lorenz curve and Gini index. The Review of Economics and Statistics, pp. 306-316, 1972.

Gelhausen, M. C. **Airport Choice in a Constraint World: Discrete Choice Models and Capacity Constraints**. Proceedings of the Air Transport Research Society World Conference 2008: pp. 1-16, 2008.

Graaff, J. de V. **Theoretical Welfare Economics**, Cambridge University Press, 1957.

Halcrow-Sinergia Consortium. **Brazil TAV Project – Volume 1 – Demand and Revenue Forecast**. 2009a, Available at <http://www.antt.gov.br/html/objects/_downloadblob.php?cod_blob=3463> Access in mar. 2nd, 2012.

Halcrow-Sinergia Consortium. **Brazil TAV Project – Volume 4 – Operations**. 2009b, Available at <http://www.antt.gov.br/html/objects/_downloadblob.php?cod_blob=3463> Access in mar. 2nd, 2012.

Harvey, G. **Airport Choice in a Multiple Airport Region**. Transportation Research A, volume 21A (6), pp. 439-449, 1987.

Hess, S., Polak, J. W. **Exploring the potential for cross-nesting structures in airport-choice analysis: A case-study of the Greater London area**. Transportation Research Part E: Logistics and Transportation Review, Volume 42, Issue 2, Pages 63-81, 2006a

________. **Airport, Airline and Access Mode Choice in the San Francisco Bay Area**. Papers in regional Science 85.4, Pages 543-567, 2006b.

Hoffman, S. D., Duncan, G. D. **Multinomial and Conditional Logit Discrete-Choice Models in Demography**. Demography, Vol. 25 No. 3, pp. 415-427, 1988.

IAC (Instituto de Aviação Civil). **Demanda Detalhada dos Aeroportos Brasileiros**. Vol. 1, 2005. Available at <http://www2.transportes.gov.br/PNLT/DVD_AA/Demanda_Detalhada_2005/Volume1_2005.pdf >. Access in aug. 23rd , 2011.

________. **Demanda Detalhada dos Aeroportos Brasileiros**, Vol. 2, 2005 available at <http://www2.transportes.gov.br/PNLT/DVD_AA/Demanda_Detalhada_2005/Volume2_2005.pdf >. Access in aug. 23rd , 2011.

IBGE (Instituto Brasileiro de Geografia e Estatística). **Contas Nacionais Trimestrais – Abril/Junho 2012**. Rio de Janeiro: IBGE, 2012

________. **Contas Nacionais Trimestrais – Abril/Junho 2011**. Rio de Janeiro: IBGE, 2011a.

________. **Sinopse do Censo Demográfico 2010**. Rio de Janeiro: IBGE, 2011b.

________. **Contas Nacionais Trimestrais – Abril/Junho 2010**. Rio de Janeiro: IBGE, 2010.

INFRAERO. **Movimento Operacional da REDE INFRAERO de Janeiro a Dezembro de 2011.** Available at <http://www.infraero.gov.br/images/stories/Estatistica/2011/Dez.pdf> Access in nov. 8th, 2012, 2011.

________.Movimento Operacional da REDE INFRAERO de Janeiro a Dezembro de 2010. Available at <http://www.infraero.gov.br/images/stories/Estatistica/2010/Dez.pdf > Access in nov. 8th, 2012, 2010

________.Movimento Operacional da REDE INFRAERO de Janeiro a Dezembro de 2009. Available at

<http://www.infraero.gov.br/images/stories/Estatistica/2009/mov_operac._1209_revisado.pdf> Access in nov. 8th, 2012, 2009

Innes, J.D.; Doucet D. H. **Effects of Access Distance and Level of Service on Airport Choice**. Journal of Transportation Engineering, volume 116 (4), pp 507-516, 1990.

McFadden. D. **Conditional Logit Analysis of Qualitative Choice Behavior**. Frontiers in Econometrics. Edited by P. Zarembka, Academic Press, 1973.

McKinsey and Company, **Estudo do Setor de Transporte Aéreo do Brasil –Relatório Consolidado**. 2009. Available at <http://www.bndes.gov.br/SiteBNDES/export/sites/default/bndes_pt/Galerias/Arquivos/empresa/pesquisa/chamada3/relatorio_consolidado.pdf>. Access in mar. 2nd, 2012.

Moreno, M.B.; Müller, C. **Escolha de Aeroportos em Região de Múltiplos Aeroportos": O caso da Grande São Paulo**. Masters Thesis – ITA – Civil Engineering Department, 2002 Available at <http://161.24.11.4/tesesdigitais/verifica_session.php?num_tese=000485762&origem=BD>. Access in mar. 2nd, 2012.

________. **The influence of ground accessibility on airport choice in São Paulo metropolitan area**. Paper presented at the 19th Dresden Conference of Traffic and Transportation Sciences, 2003. Available at < http://vwitme011.vkw.tu-dresden.de/TrafficForum/vwt_2003/beitraege/VWT19proceedings_contribution_93.1-93.23.pdf>. Access in mar. 2nd, 2012.

Pels, E; Nijkamp, P; Ritveld, P. **Airport choice in a multiple airport region: an empirical analysis for the San Francisco Bay Area**. Discussion paper 98-041/3, Tinbergen Institute, Holland, 1998.

Rocha, G. C. **Ensaios sobre a Demanda do Transporte Aéreo Regional**. Journal of Transport Literature, vol. 4, n. 1, pp. 114-133, 2010.

Sen, A. **Real National Income**. The Review of Economic Studies, Vol. 4, No. 1, pp-19-39, Oxford University Press, 1982.

Skinner Jr., R. E., **Airport Choice: An Empirical Study**, Transportation Engineering Journal of ASCE, volume 102 (TE4), November, pp. 871-882, 1976.

Soares, S. **O ritmo de queda na desigualdade no Brasil é adequado?: evidências do contexto histórico e internacional.** Text for Discussion No. 1339. Brasília: IPEA, Instituto de Pesquisa Econômica Aplicada, 2008.

Stone, M.; Rasp, J. **The Assessment of Predictive Accuracy and Model Overfitting: an Alternative Approach.** Journal of Business Finance & Accounting, 20: 125–131,1993.

Windle, R.; Dresner M. **Airport Choice in Multiple Airport Regions**. Journal of Transportation Engineering, July/August, pp 332-337, 1995.

Vovsha, P. **Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area**. Transportation Research Record: Journal of the Transportation Research Board, Volume 1607, pp 6-15, 1997.

Annex A

Interview forms used by FIPE (2009) at Brazilian airports to gather passenger data – English Version fipe

CARACTERIZAÇÃO E DIMENSIONAMENTO DA MATRIZ ORIGEM DESTINO DO TRANSPORTE AÉREO NO BRASIL

Aeroporto:________________________________ **Sigla:** __________ **Cód. do Questionário:** ______________

Entrevistador: __________________________ **Cód.:** _________ **Data:** ____/____/09 **Hora:**_____:_____

Good Morning/Afternoon/Evening, my name is_____. I'm a researcher from FIPE and I'm conducting a survey for BNDES that will assist in the future investments in airport infrastructure in Brazil. Can we talk for a few minutes?

F1. Are you beginning this part of the trip in this airport or are you coming from another one?
1. This airport (*vá para Q1*) 2. Another airport. **Which one?**

Airport:____________________ City:_________________________ State: _____ Country: ____________

✓ *Encerre a entrevista se vier de um outro aeroporto onde há pesquisa.*

A. IDA OU VOLTA

1. In which city do you LIVE?

City:____ State: _____ Country: ____________________

2. IF YOU LIVE IN BRAZIL, which is the zip code of your residence? ____________________ ☐ NSR

3. In this PART OF THE TRIP, you are: 1. Going 2. Returning 3. Going-round 4. One way trip

B. CRONOLOGIA DA VIAGEM

4. In which CITY DID YOU START this part of the trip? Considere o município inicial mesmo que ele não seja o município onde fica o aeroporto do primeiro embarque.

City:____ State: ______ Country: _________________

✓ *Se o aeroporto inicial é estrangeiro, pule para Q9.*

5. In this city, which PLACE did you leave from to arrive in this airport?
1. Own home/Friends and relatives home 3. Hotel or similar 5. Place of event
2. Work 4. Place where you study (high school, college, etc) 6. Other. Which? _________

➔ *Q4 = SÃO PAULO, BRASÍLIA, BELO HORIZONTE ou RIO DE JANEIRO, pergunte:*

6. Which is the ZIP CODE of this place? ________________________ ☐ Já informou ☐ NSR

7. Which is the NEIGHBORHOOD? ____ ☐ NSR

8. Which MEANS OF TRANSPORTATION did you use on the way to the airport? Informar na seqüência de uso. Considerar inclusive meios de transporte utilizados dentro do mesmo município. Não considerar avião. *(quadro branco 1)*

Means of transportation 1° ______ 2° ______ 3° ______ 4° ______
1.Cab 2.Own car 3.Lift 4.Motorcycle 5.Rented car 6.Car owned by the company 7.Transfer 8.Subway 9.Train 10.Regular bus 11.Airline bus 12.Intercity, interstate o international bus 13.Tour bus 14.Ferryboat, ship or flatboat 15.Other: ___________

8.1 HOW LONG did this part for the travel take?

Tempo total - Days: _______ Hours: ________ Minutes: ________ ☐ NSR

8.2 How much HAVE YOU SPENT on this part of the trip?

Ammount _______________,00 Currency:________ ☐ NSR

8.3 ➔ ***SOMENTE SE*** *o último meio de transporte for carro próprio (2), carona (3) ou moto (4):*

Did you or the person who brought you use the PARKING FACILITIES?

1. No 2. Yes. **For how long?** Days: ______ Hours: ________ Minutes: ________ ☐ NSR

➔ AS TO THE INITIAL AIRPORT in BRAZIL:

9. **What time did you arrive in the airport?** Time: ______:______ Date: _____/_____ ☐ NSR

10. How did you CHECK-IN your flight? *(resp. única)*
1. Balcony 2. Internet 3. Mobile 4. Totem 5. Dispatcher 6. Didn't

11. Did you dispatch your luggage? 1. Yes 2. No

12. Which is the FLIGHT NUMBER and the AIRLINE of the FIRST FLIGHT? Considerar o primeiro aeroporto, no Brasil ou no exterior.

Airline: ____________________ Flight Number: _______________

13. Which is the DEPARTURE TIME of the flight?
Time: _______:_______ Date: ______/______ ☐ NSR

14. Does this flight have any CONNECTIONS or AERO STOPS? Conexões envolvem troca de avião. Escala é a situação em que o avião aterrisa em alguma cidade SEM que o passageiro tenha que sair da aeronave para continuar a viagem.

→ 1. No

14.1. Which is the AIRPORT OF FINAL DESTINATION of the trip?
Airport:_________________________ City:_____________________ State: ______ Country: ____________

14.2. Which is the flight ARRIVAL TIME at destination?
Time: _______:_______ Date: ______/______ ☐ NSR
(pule para Q18)

→ 2. Sim. **Could you inform me the connections and aero stops in chronological order?**

15. Which is the first AIRPORT OF THE CONNECTION OR AERO STOP?

Airport:_________________________ City:_____________________ State: ______ Country: ____________

16. What is the ARRIVAL TIME of the flight at this airport?

Time: _______:_______ Date: ______/______ ☐ NSR

17. Would it be connection or aero stop? ☐ Aero stop ☐ Connection

What is the DEPARTURE time?	Which AIRLINE and FLIGHT?	Which AIRPORT?	What is the ARRIVAL time?	Is it the FINAL airport?
_____:_____ Date: ____/____ ☐ NSR	Airline: ___________ Flight: ___________	Airport: ___________ City: ___________ State: ________ Country: ___________	_____:_____ Date: ____/____ ☐ NSR	☐ Yes (pule p/ Q18) No: ☐ Connection ☐ Aero stop
_____:_____ Date: ____/____ ☐ NSR	Airline: ___________ Flight: ___________	Airport: ___________ City: ___________ State: ________ Country: ___________	_____:_____ Date: ____/____ ☐ NSR	☐ Yes (pule p/ Q18) No: ☐ Connection ☐ Aero stop
_____:_____ Date: ____/____ ☐ NSR	Airline: ___________ Flight: ___________	Airport: ___________ City: ___________ State: ________ Country: ___________	_____:_____ Date: ____/____ ☐ NSR	☐ Yes (pule p/ Q18) No: ☐ Connection ☐ Aero stop
_____:_____ Date: ____/____ ☐ NSR	Airline: ___________ Flight: ___________	Airport: ___________ City: ___________ State: ________ Country: ___________	_____:_____ Date: ____/____ ☐ NSR	☐ Yes (pule p/ Q18) No: ☐ Connection ☐ Aero stop
_____:_____ Date: ____/____ ☐ NSR	Airline: ___________ Flight: ___________	Airport: ___________ City: ___________ State: ________ Country: ___________	_____:_____ Date: ____/____ ☐ NSR	☐ Yes (pule p/ Q18) No: ☐ Connection ☐ Aero stop

18. Which city is the FINAL DESTINATION of this part of the trip? Não necessariamente o município final é o mesmo do aeroporto de desembarque.

City:____ State: _______Country: __________________

➔ *Q18 = SÃO PAULO, BRASÍLIA, BELO HORIZONTE ou RIO DE JANEIRO, pergunte:*

19. Which is the ZIP CODE of the destination in that city?______________________ ☐ Já informou ☐ NSR

20. Which is the NEIGHBORHOOD? ___________________________________ ☐ NSR

✓ *Se o aeroporto final é estrangeiro, pule para Q22.*

21. Which MEANS OF TRANSPORTATION do you intent to use on the way from the airport to the final destination? *(quadro branco 1)*

Means of transportation 1° ______ 2° ______ 3° ______ 4° ______

1.Cab 2.Own car 3.Lift 4.Motorcycle 5.Rented car 6.Car owned by the company 7.Transfer 8.Subway 9.Train 10.Regular bus 11.Airline bus 12.Intercity, interstate o international bus 13.Tour bus 14.Ferryboat, ship or flatboat 15.Other: ____________

21.1 HOW LONG do you think that this part of the trip will take?

Tempo total - Days: _______ Hours: ________ Minutes: ________ ☐ NSR

21.2 How much do you think that this part of the trip WILL COST?

Ammount ____________________,00 Currency:__________ ☐ NSR

21.3 *SOMENTE SE* *o primeiro meio de transporte for carro próprio (2), carona (3) ou moto (4)*:

Are you, or the person who will pick you up is going to use the PARKING FACILITIES of the LANDING AIRPORT?

1. No 2. Yes. **For how long?** Days: _______ Hours: ________ Minutes: ________ ☐ NSR

C. INFORMAÇÕES DA VIAGEM E DA COMPRA

22. What is the MAIN REASON of this trip to Brazil? *(resp. única)*

1. Business or professional events
2. Leisure
3. Visit to friends or relatives
4. Study purposes
5. Social, cultural or sports events
6. Other. Which? _________________________

23. HOW LONG AHEAD did you buy your airline ticket? Days: _____ Hours: ______ Minutes: ______ ☐ NSR

24. WHO HAS PAID for your airline ticket?? *(permitir múltiplas respostas)*

1. Yourself
2. Place where you work
3. Mileage *(pule para Q26)*
4. Friend or relative
5. Other: ______________

25. How much HAVE YOU SPENT IN YOUR AIRLINE TICKET for this part of the trip? (individual value, only yours, including all taxes)? Confirmar no bilhete ou e-ticket se necessário.

Ammount _______________________,00 Currency:________ 1. Just going or jus coming 2. Round trip ☐ NSR

26. DID YOU BUY ANYTHING during your permanence at this airport?

1. No 2. Yes. **How much have you spent?** Ammount _______________________,00 Currency:________

D. PREFERÊNCIAS E OPINIÕES DO ENTREVISTADO

✓ *Se o aeroporto inicial é estrangeiro, pule para Q29.*

27. Which are the three MAIN REASONS for choosing the INITIAL AIRPORT? *(quadro vermelho 2)*

1. The airport has a direct flight to the destination.
2. It is the closest airport to the origin.
3. The access to the airport is cheaper.
4. The favorite airline uses this airport.
5. It is the airport that offers the best prices of airline tickets.
6. The airport offers the best days/times of flights.
7. It is the only option.
8. I didn't choose
9. Other: __________________

28. Is the initial airport THE BEST BOARDING OPTION in regard to other Brazilian airports?

1. Yes
2. No, it would be BETTER in: Airport:__________________________ City:_______________________ State: _____

✓ *Se o aeroporto final é estrangeiro, pule para Q31.*

29. Which are the three MAIN REASONS for choosing the FINAL AIRPORT? *(quadro vermelho 2)*

1. The airport has a direct flight to the destination.
2. It is the closest airport to the destination.
3. The access to the airport is cheaper.
4. The favorite airline uses this airport.
5. It is the airport that offers the best prices of airline tickets.
6. The airport offers the best days/times of flights.
7. It is the only option.
8. I didn't choose
9. Other: __________________

31. Which are the three main reasons for CHOOSING THE FLIGHTS of this part of the travel?
(quadro laranja 3)

1. Price
2. Flight duration
3. Better Day/ Time option
4. Fewer connections/aero stops
5. My favorite flight was overbooked
6. That was the only available flight
7. Airline punctuality
8. Mileage Programs
9. Airline services (board services, buses etc)
10. I didn't choose
11. Others: ____________

32. Would you prefer if this flight were available at any OTHER DATE?

1. No 2. Yes. **What would be the preferred date?** Date: _____/_____/______

33. Would you prefer if this flight were available at any OTHER TIME?

1. No 2. Yes. **What would be the preferred time?** Time: ______:______

34. Including this trip, HOW MANY TIMES did you do THIS PART OF THE TRIP BY AIRPLANE in the last 12 months? ________

E. VIAGENS COM OUTROS MEIOS DE TRANSPORTE

35. Have you made this part of the trip USING ANOTHER MEAN OF TRANSPORTATION in the last 12 months?

1. No *(pule para bloco F)* 2. Yes. **HOW MANY times?** ___________

36. WHICH OTHER kind of TRANSPORTATION do you use in this part of the trip instead of airplanes?
(quadro branco 1)(resp. única)

1. Cab
2. Own car
3. Lift
4. Motorcycle
5. Rented car
6. Car owned by the company
7. Transfer
8. Subway
9. Train
10. Regular bus
11. Airline bus
12. Intercity, interstate o international bus
13. Tour bus
14. Ferryboat, ship or flatboat
15. Other: ____________

37. WHY did you choose this kind of transportation instead of airplane? *(quadro amarelo 4)(resp. única)*

1. Prices
2. Travelling time
3. Security
4. Date/time
5. Flights delays
6. Difficulty to access the airport
7. Flights overbooking
8. Mobility on the destination
9. I didn't choose
10. Other: ____________

F. DADOS DO ENTREVISTADO

38. Including this one, how many times have you MADE THIS PART OF THE TRIP using AIRPLANES in the last 12 months to ANY DESTINATION? __________ *(ida e volta contam separadamente)*

39. In which COUNTRY DID YOU LIVE most part of the time in last 12 months?

1. Brazil 2. Other: ___________________

40. What is your CITIZENSHIP? 1. Brazilian 2. Other: ___________________

41. (não perguntar) GÊNERO: 1. Masculino 2. Feminino

42. What is your AGE? ________

43. What is your family monthly INCOME? _____________________,00 Currency:________ *(pule p/ Q 44)*

☐ Doesn't know/ didn't want to inform ➔ *insista com o quadro azul 5* ☐ Doesn't have family income *(pule para 45)*

43.1. Class of family monthly income: *(quadro azul 5)*

1. Até R$ 930,00
2. De R$ 930,01 até R$ 2.325,00
3. De R$ 2.325,01 até R$ 4.650,00
4. De R$ 4.650,01 até R$ 6.975,00
5. De R$ 6.975,01 até R$ 9.300,00
6. De R$ 9.300,01 até R$ 13.950,00
7. De R$ 13.950,01 até R$ 23.251,00
8. Acima de R$ 23.251,01

44. INCLUDING YOURSELF, how many people depend on this income? _______

45. Which is your MAIN PROFESSIONAL OCCUPATION at this time? *(quadro verde 6)*

1. Hireling
2. Civil Servant
3. Liberal Professional
4. Entrepreneur
5. Student, trainee or scholarship student
6. Retired or pensioner
7. Housewife
8. Unemployed
9. Other: _______________

Supervisor: _________________________ (________) **Checador:**____________________ (________)

Annex B

Number of weekly flights originating from the São Paulo Area airports as of 08/29/2009 (ANAC, 2009)

AIRPORTS	FDW
ORIGIN - GRU	**1847**
Afonso Pena	91
AMSTERDAM SCHIPOL	7
ASUNCION	14
AUGUSTO SEVERO	37
Bahia - Jorge Amado	7
BOGATA	7
BUENOS AIRES MINISTRO PISTA	118
Campo Grande	27
CARACAS SIMON BOLIVAR	7
Cataratas	28
CHICAGO O' HARE	7
CIUDAD DEL ESTE	7
DAKAR YOFF	3
DALLAS-FT WORTH	7
Deputado Luís Eduardo Magalhães	140
DUBAI	7
Eduardo Gomes	41
FRANKFURT	14
G BUSH INTERCONTINENTAL	7
Galeão - Antônio Carlos Jobim	128
Goiabeiras	21
Guararapes Gilberto Freyre	100
Hercílio Luz	55
JOHANNESBURG JAN SMUTS	7
Lauro Kurtz	6
Leite Lopes	33
LIMA-CALLAO	24
LISBON	11
LONDON HEATHROW	14
LOS ANGELES	3
LUANDA 4 DE FEVEREIRO	3
MADRID BARAJAS	21
Marechal Cunha Machado	7
Marechal Rondon	14
MEXICO CITY	12
Miami	30
MILANO MALPENSA	10
Ministro Victor Konder	7
MONTEVIDEO CARRASCO	31
MUENCHEN	5
NEW YORK	30
NEWARK	7
OPORTO	3
ORLANDO INTL	7

PANAMA CITY	14
PARIS-CHARLES-DE-GAULLE	28
Pinto Martins	48
Porto Seguro	10
Pres. Juscelino Kubitschek	98
Presidente Castro Pinto	14
Regional de Maringá - Sílvio Name Júnior	5
ROMA FIUMICINO	6
Salgado Filho	123
Santa Genoveva	28
SANTA MARIA	7
SANTIAGO	44
Santos Dumont	67
São José do Rio Preto	1
Tancredo Neves	76
Ten. - Cel. Av. César Bombonato	18
THE HARTSFIELD ATLANTA	7
TORONTO LESTER	7
Uberaba	6
Usiminas	6
Val de Cans	21
VIRU VIRU	8
WASHINGTON DULLES	7
Zumbi dos Palmares	27
ZURICH	6
ORIGIN - VCP	**459**
Afonso Pena	61
Campo Grande	13
Deputado Luís Eduardo Magalhães	44
Eduardo Gomes	14
Galeão - Antônio Carlos Jobim	45
Goiabeiras	20
Guararapes Gilberto Freyre	14
Hercílio Luz	7
Londrina	1
Ministro Victor Konder	21
Pampulha - Carlos Drummond de Andrade	12
Pinto Martins	14
Pres. Juscelino Kubitschek	32
Regional de Maringá - Sílvio Name Júnior	14
Salgado Filho	35
Santos Dumont	45
Tancredo Neves	60
Zumbi dos Palmares	7
ORIGIN - CGH	**989**
Afonso Pena	120

Bahia - Jorge Amado	14
BAURU/AREALVA	34
Campo dos Bugres	12
Campo Grande	33
Deputado Luís Eduardo Magalhães	13
ESTADUAL DARIO GUARITA	11
Francisco de Assis	22
Galeão - Antônio Carlos Jobim	61
Goiabeiras	55
Hercílio Luz	75
Lauro Carneiro de Loyola	39
Leite Lopes	20
Londrina	27
Marechal Rondon	13
Marília	24
Ministro Victor Konder	40
Pres. Juscelino Kubitschek	7
Presidente Prudente	13
Regional de Maringá - Sílvio Name Júnior	18
Salgado Filho	96
Santa Genoveva	50
Santos Dumont	440
São José do Rio Preto	21
Tancredo Neves	137
Ten. - Cel. Av. César Bombonato	34
SUM OF FREQUENCIES	**3295**

Annex C

Stata do-files for fitting conditional logit models

File for conditional logit for the São Paulo and Campinas Metropolitan Areas:

```
clear all
capture log close
set more off
use c:/metro.dta
label define air 1 "gru" 2 "cgh" 3 "vcp"
label define pur 1 "business" 2 "leisure
label define type 1 "intl" 0 "dom
label values airport air
label values purpose pur
label values intl type
gen id=_n
reshape long d e t f, i(id) j (airchoice gru cgh vcp) string
asclogit d e t f, case(id) alternatives(airchoice)
estat mfx
predict prob, pr
log close
```

File for conditional logit for the São Paulo state outside the São Paulo and Campinas Metropolitan Areas:

```
clear all
capture log close
set more off
use c:/int.dta
label define air 1 "gru" 2 "cgh" 3 "vcp"
label define pur 1 "business" 2 "leisure
label define type 1 "intl" 0 "dom
label values airport air
label values purpose pur
label values intl type
gen id=_n
reshape long d e t f, i(id) j (airchoice gru cgh vcp) string
asclogit d e t f, case(id) alternatives(airchoice)
estat mfx
predict prob, pr
log close
```

<table>
<tr><td colspan="4">FOLHA DE REGISTRO DO DOCUMENTO</td></tr>
<tr><td>1. CLASSIFICAÇÃO/TIPO
TC</td><td>2. DATA
21 de novembro de 2012</td><td>3. REGISTRO N°
DCTA/ITA/TC-107/2012</td><td>4. N° DE PÁGINAS
103</td></tr>
<tr><td colspan="4">5. TÍTULO E SUBTÍTULO:
Evaluation of long-term demand for the Viracopos Airport under accessibility hypothesis with or without high-speed railway.</td></tr>
<tr><td colspan="4">6. AUTOR(ES):
Renan Rios Diniz</td></tr>
<tr><td colspan="4">7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES):
Instituto Tecnológico de Aeronáutica – ITA</td></tr>
<tr><td colspan="4">8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR:
Viracopos Airport, TAV, Demand forecast, Discrete choice, Conditional logit</td></tr>
<tr><td colspan="4">9.PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO:
Aeroportos; Transporte de passageiros; Transporte ferroviário; Estimativas; Planejamento de aeroportos; Demanda (Economia); Previsão; Engenharia de transportes.</td></tr>
<tr><td colspan="4">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 2012.</td></tr>
<tr><td colspan="4">11. RESUMO:
This graduation thesis seeks to understand the impact of building or not building the TAV – a high speed railway linking Rio de Janeiro, São Paulo and Campinas – on the São Paulo Area airports, especially the Viracopos airport, in Campinas. This is done through the definition of a demand forecast model for the airports the Campinas and São Paulo Metropolitan areas, and through the development of a discrete choice model for the Viracopos, Guarulhos and Congonhas airports, in the same region. The demand forecast model is based on a review of the estimates done by the company McKinsey (2009). Such reviews incorporate an adjustment for changes in income distribution based on a metric developed by Sen (1982), as well as adjustments to GDP, yield and the passenger traffic in 2009. The discrete choice model developed for the airports, supported by literature review, is a conditional logit model with segments for the São Paulo and Campinas Metropolitan Areas, taken together, and the rest of the São Paulo State. Frequency and access time are relevant variables in the first case, and only frequency in the second, creating, regarding the influence of passenger experience with airports with a previous work by Moreno and Müller (2003). The discrete choice model is integrated to the demand forecast and used to discuss the impacts on the use of each airport in 2030 showing a possibility of quick saturation of the airports and a need for improving access to Viracopos. The thesis concludes with policy and research recommendations based on the results.</td></tr>
<tr><td colspan="4">12. GRAU DE SIGILO:
(X) OSTENSIVO ( ) RESERVADO ( ) CONFIDENCIAL ( ) SECRETO</td></tr>
</table>