Market share modeling in a newly liberalized low-demand market

The case study of the Azorean archipelago

Britta Wilken

Aerospace Engineering - Air Transport and Operations
Confidential

Market share modeling in a newly liberalized low-demand market

The case study of the Azorean archipelago

by

Britta Wilken

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday July 22^{nd}, 2016 at 14:30 o'clock.

Student number:	4056426	
Project duration:	April, 2015 – July, 2016	
Thesis committee:	Prof. dr. R. Curran,	TU Delft
	Dr. B. F. Lopes dos Santos,	TU Delft, supervisor
	Dr. M. Kroesen,	TU Delft
	Dr. E. van Kampen,	TU Delft

TU Delft

Preface

The idea of collaborating with the Azorean airline SATA for my MSc thesis emerged during my internship on the Azores. I was helping to find a flight schedule with which SATA was able to compete against the low-cost carriers (LCC's). At that point it was not clear what their schedule would be or with what range of ticket prices they would enter the market. I could feel how much impact competitors have on the operations of an airline.

I found this to be a very interesting environment to write my thesis about. I believed, that if I could capture the impact of liberalization on that market, this information would be beneficial for SATA when the LCC's will start operations to and from Terceira (another Azorean island which is also liberalized but the LCC's did not choose to fly there yet) or other comparable markets.

This is how this MSc thesis came into being and it would not have been possible without the guidance of my supervisor Bruno Santos from Delft University of Technology. He supported my in every possible way during my project and because of his input, after every meeting I felt like I was a step closer to my graduation.

Additionally, I would like to thank Maarten Kroesen from the faculty of Technology, Policy and Management at Delft University of Technology. He led me in the right direction at the beginning of the process, when I was struggling with the statistical side of the project. His critical insight during key meetings improved my work and his attitude always lightened up the atmosphere.

Next to the support from the university, I would like to thank my colleagues from the Azorean airline SATA. It was a pleasure working with you during my internship. Especially, I would like to thank Sandro Raposo and Filipe Raposo for all the effort they made to provide me with the required input data. Additionally, it was very valuable that they made me look at the project from an airline perspective. Thank you very much for everything both of you have done for me.

Last but not least, I would like to thank everyone close to my heart for their mental support during this project. Thanks for the distraction, thanks for listening to my complaints, thanks for nice dinners, and the contribution you made to my thesis.

In Gedanken bin ich bei dir, Oma.

Britta Wilken
Delft, March 2016

Summary

In the last decades the airline industry experienced a tendency towards airspace deregulation. When airspace is liberalized, the airlines in the market have to position themselves newly, since new airlines are expected to enter the market. Decisions have to be made about ticket price, scheduling, and service even though it is uncertain what kind of product the competition will offer and how the passenger will adjust their booking behavior. For the airline it is important to understand which airline attributes have significant effect on the market share. If this is know, these attributes can be the focus when making strategic and tactical decisions. More competition asks for optimal revenue management and positioning of the flights so that no market shares are lost.

The goal of this research is to create a market share model that is able to detect the airline attributes that show significant impact on market share in the markets before and after liberalization. This way, conclusions can be drawn about the changes in the market due to liberalization. Furthermore, the market share model should be able to compare the different effects of the airline attributes on market share between each other.

To develop such a market share model, reference data is required, which means that the model needs to be applied to a case study. For this research, the Portuguese Azores archipelago was taken as a case study, since the airspace was liberalized recently in 2015. This research was performed as a MSc thesis at the Delft University of Technology and in collaboration with the Azorean airline SATA. The market share model can potentially help SATA to position their airline against two low-cost carriers (LCC's): easyjet and Ryanair, which have entered the market between Ponta Delgada (main city in the archipelago) and Lisbon as well as Porto. The route between Terceira (one of the nine Azorean islands) and the Portuguese mainland is also liberalized but no airline has entered the market yet. The market share model could potentially be applied to this market (and other markets) as well, since it has the same main characteristic: low-demand.

Literature

Next to the potential gain for the industry, benefits can also be formulated from the academic point of view. Adler [1] investigated the impact of liberalization on air traffic and showed that due to liberalization the passenger traffic will increase but the ticket price and therefore yield of the airline will decrease. This is due to an increase in competition and therefore more pressure is present on the pricing and service quality of an airline [14].

Fu et al. [17] stated that also economic growth is triggered by liberalization due to employment opportunities, trade promotion and better transport and logistic services. O'Connell and Williams [30] concluded from their analysis, that legacy carriers are not able to match the fares of the low-cost carrier due to, amongst other factors, inefficient operating practices. To cut some costs, legacy airlines tend to optimize their networks after liberalizations according to Fu et al. [17]. This means that the legacy carriers try to move the network more into the direction of a 'hub-and-spoke' (HS) network to reach

more passengers.

To built the model that simulates the liberalization situation in a low-demand environment an appropriate methodology needs to be found. Market share can be calculated using different methods such as: game theory [21], fuzzy logit [20], or aggregate-level Markovian type model [43]. However, the model with the most potential to determine market share was found to be the logit model, as addressed by Wei and Hansen [49] and Adler et al. [3], who also linked liberalization to market share modeling.However, because no binary data was available, the appropriate methodology is the generalized linear model (GLM) with logit link. The method is able to process fractional data when introducing the Bernoulli log-likelihood estimation which was proposed by Papke and Wooldridge [32].

Research question

The research question that is needed to be answered during the project was formulated as follows:

> *What is the short-term impact of liberalization on airline market share in a newly liberalized low-demand market?*

To be able to answer this question, it was broken down into several sub-questions. These focused on finding the significant airline attributes and their influence on the low-demand market before and after liberalization. In the end, an equation needs to be constructed which will enable one to forecast the market share of the airlines involved in the market after liberalization.

This research question is relevant for the industry, as well as from the academic point of view. The market share model that is developed can help airlines that are facing liberalization to make decisions on how to place their airline in the market. This goal can be achieved when knowing which airline attributes have significant impact on market share and how this is affected by liberalization. Adding to the body of knowledge, the generalized linear model with logit link has not been applied before to airline market share. Papke and Wooldridge [32] applied the GLM to a participation rate of a pension plan which was given as a fraction. Also, the logit model used in several papers ([3], [37], [50]) to draw conclusions about influence of airline attributes on market share. However, a market share model has not been constructed with the GLM method and applied to the liberalization context. This in a novelty and the contribution of this research.

Model

The GLM was selected as the appropriate methodology for the market share model. Since the response variable is the market share, a logit link function and binomial family was chosen. Also, robust standard errors had to be introduced because of the independent variable is not necessary identically distributed [40]. The statistic software STATA® was selected for the calculations because it was able to used the GLM with these characteristics.

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \tag{1}$$

$$marketshare = \frac{exp(\eta)}{1 + exp(\eta)} \tag{2}$$

Airline attributes take over the function as independent variables that are denoted with x in the score η in Equation 1. They were selected based on literature study and their expectancy to have impact on market share in the proposed situations. These attributes are: *ticket price the day prior departure ($price_{day}$) and two months prior departure ($price_{2m}$), frequency per day ($freq_{day}$) and per week ($freq_{week}$), time points ($time\_points$) and on-time performance (on_time).* An additional market share model is developed that included monthly dummy variables as well. The independent variables represent the airline attributes of SATA in relation to its competitors.

The β-coefficients in Equation 1, are the unknowns that need to be determined by the GLM calculation. With this output, the market share can be calculated using Equation 2. For the variables, calibration data was needed for the situation before and after liberalization. Passenger data was available for the Azorean airline SATA and the Portuguese airline TAP. These included all operational data of each flight. To calculate the market share of the LCC's, the airport report of Ponta Delgada airport was used. Furthermore, the price data was partly received by SATA and partly collected in the online booking system.

The market share model was developed for three different sets of independent variables for all the markets and situations. First, all airline-related independent variables are included, then, the variable that represents $freq_{day}$ is removed from the model because of high correlation and in the third model, all airline-related variables are expanded by monthly dummy variables.

The β-coefficients were calculated for two different situations: before and after liberalization for the market between Ponta Delgada and Lisbon. For the market between Ponta Delgada and Porto, a market share model was developed for the situation after liberalization only, because before liberalization, SATA had a monopoly position on this route.

Results

The β-coefficients were calculated for three sets of independent variables, two situations and two markets, as well as the combined market of these two, as indicated above. In all market share models the β-coefficient of the independent variable that represents $freq_{day}$ is significant (if included in the model) and takes over a significant share of the impact on market share. The variable showed high correlation with market share which makes sense because if the frequency is higher, more seats are offered in a route and a higher market share can be achieved.

Next to that, the constant, that is indicated as β_0 in Equation 1, is significant in almost all market share models. It represents all effects that have an influence on market share that are not represented by the independent variables. The constant and $freq_{day}$ variable were found to have the biggest impact on market share in all the models.

Next to $freq_{day}$ and the constant, the significant independent variables vary. For the combined market between Ponta Delgada and Lisbon, as well as Porto after liberalization, only $freq_{day}$ and the constant are significant even when adding monthly dummies. When $freq_{day}$ is removed from the set of independent variables, $price_{2m}$ and $time\_points$ are significant and the constant is included which all three have comparable impact on market share.

For the market between Ponta Delgada and Lisbon after liberalization, $price_{day}$ is the additional significant variable for the first set of independent variables. For the second set, $time\_points$ is significant instead of $freq_{day}$, and for the last market share model, all monthly dummy variables show increasing impact on market share approaching the summer months.

Before liberalization, only $freq_{day}$ and the constant have impact on market share for the same market. When $freq_{day}$ is removed, $price_{2m}$ and $freq_{week}$ (+constant) are significant which are independent variables that did not show effect in the situation after liberalization.

The market between Ponta Delgada and Porto showed some unexpected results. When all airline-related variables are included, $freq_{day}$ and $freq_{week}$ are significant but the constant is not. Also, the β-coefficient of $freq_{week}$ is negative. A higher $freq_{week}$ is expected to trigger higher market shares which means that the β-coefficient is supposed to be positive. The same phenomenon was found in the market share model where $freq_{day}$ is deleted. In this model, only $freq_{week}$ is significant and its β-coefficient is also negative. The reason for this result lies in the fact that partucularly for this model, many observations were deleted from the calibration data. A limited set of data can lead to unexpected results.

The market share model with monthly dummies for the market between Ponta Delgada and Porto after liberalization showed satisfying results. $Price_{day}$ and $freq_{day}$ as well as the monthly dummies $june$ and $july$ are significant. Approaching the summer months, the market share of SATA decreased, which could be seen from the β-coefficients that become more negative. A reason for this could be that for the flights in spring time, tickets were already sold before the entrance of the low-cost carriers were announced.

Sensitivity analysis and validation

In the sensitivity analysis, the effect of the maximum and the minimum value of an independent variable on market share is determined while all other variables are set to the mean value. The biggest changes in market share are triggered by the independent variable $freq_{day}$. The highest change value for $freq_{day}$ was found in the return market between Ponta Delgada and Lisbon after liberalization for all airline-related variables and is equal to 24.3%. For the same market circumstances, the lowest value is equal to -22.6% and was calculated for the combined market. The remaining significant variable trigger market share changes of under 8%.

Comparing the situation before and after liberalization, it can be seen that the changes in market share are significantly higher after liberalization. The reason for this lies in the fact that the constant has more impact on the market share in the situation before liberalization and is therefore less prone to the changes in the independent variable.

For the validation, the calculated β-coefficients and the values for the independent variables are inserted in Equation 1 that can be used to fulfill the purpose to estimating the market share in the validation month September 2015. The validation could be performed for the market share models after liberalization because for the situation before liberalization, no validation data was available. Additionally, the market share with monthly dummy variables was also not validated.

The market share models for the combined market and the market between Ponta Delgada post-liberalization performed well in the validation and are supposed to be suited for estimating the market share of the future. The estimated market shares of the market share models with all airline-related independent variables for the combined route and the route between Ponta Delgada is less than 1 percentage point off for the validation month September 2015. When $freq_{day}$ is removed from the model, the fit of the model decreases which can also be seen in the validation. The difference between actual and prospected market share are 8.5 and 3 percentage points for the combined and Ponta Delgada to Lisbon market respectively.

The market share model for the market between Ponta Delgada and Porto did not perform as well as the models described before. Part of the β-coefficients in the model did not have the expected sign. The difference between prospected and actual market share reaches up to 12.5 percentage points for the market share model with all airline-related independent variables and is equal to 2.5 percentage points for the model where $freq_{day}$ is removed.

Conclusion and reflection

The research has shown that the GLM with logit link function is an applicable methodology choice for a market share model. It was possible to identify the significant independent variables that represent the airline attributes in each market. The airline attribute that has the highest effect on market share was found to be frequency per day. This independent variable was significant in all market share models when included. Also, the constant was found to have an important influence on market share in almost all models.

When comparing the situation before and after liberalization for the return market between Ponta Delgada and Lisbon, it can be concluded that $freq_{day}$ has 10 percentage points higher impact on market share after liberalization, whereas the impact of the constant decreased by 20 percentage points due to liberalization. For the market share model with all airline-related independent variables, no other variable was found to be significant before liberalization, and after liberalization $price_{day}$ achieved an impact of 10% on market share. This higher importance of the ticket price could also be found in the results of the market share model where $freq_{day}$ was disregarded. This finding supports the conclusion of Adler [1], that liberalization triggers more competition and results in higher pressure on price.

Other conclusions that can be drawn from the market share model without $freq_{day}$ are: either $freq_{week}$ or $time\_points$ as well as one of the the ticket price variables take over the position of $freq_{day}$. Finally, the monthly dummy variables were found to have a positive impact on the market share on the route between Ponta Delgada and Lisbon towards the summer, a negative effect on the route between Ponta Delgada and Porto, and no effect on the combined market. This means that on the route to Porto, many seats in the spring were sold before the announcement of the liberalization and on the route to Lisbon, seasonality has a major effect.

The main limitation of the research is the available revealed preference data set. The data available for the independent variables was limited and shows little variation. The problem with the variation is that small changes in the independent variable trigger high changes in the market share model. Therefore, the model is less stable. Also, having more observations of higher quality is expected to have an effect on the calibration and validation of the model. Additionally, if there are observations available over a broader time span, the market share model could not only detect short-term changes, but even effect on the long term could be found.

Other improvements of the model could be the investigation of the effect of stop-overs and travel time on market share. In this case study, only direct flights were considered so it is recommended to implement these effects when applying the market share model in other recently liberalized low-demand markets.

Contents

List of Figures xv

List of Tables xvii

List of Acronyms xix

List of Symbols xxi

1 Introduction 1

2 Literature study 3

2.1 Liberalized situation and influence of low-cost carriers . . . 3
2.2 Methodologies and impact of airline attributes on market share . . . 4
2.2.1 Quality of service index . . . 4
2.2.2 Logit model . . . 5
2.2.3 Binary logit model . . . 6
2.2.4 Multinomial logit model . . . 6
2.2.5 Nested logit model . . . 7
2.3 Generalized linear model. . . . 8
2.4 Game theory . . . 9
2.5 Other methods . . . 10
2.6 Methodology discussion and selection . . . 11

3 Research Methodology 13

3.1 Research question and objectives . . . 13
3.2 Conceptual model . . . 14
3.3 Relevance . . . 15

4 Market share model 17

4.1 Structure of the market share model . . . 17
4.2 Quasi-maximum likelihood estimation of a fractional response variable . . . 18
4.3 Robust Standard Error . . . 19
4.4 Wald test and p-value . . . 20
4.5 Correlation . . . 21
4.6 Goodness of fit . . . 21
4.6.1 Pseudo R^2 . . . 21
4.6.2 Deviance . . . 22
4.6.3 Pearson's goodness of fit . . . 22
4.7 Impact of the independent variable on market share . . . 22
4.8 Concluding remarks of the market share model . . . 23

5 Case Study **25**
5.1 General description . . . 25
5.1.1 Public Service Obligations before liberalization . . . 26
5.1.2 After liberalization . . . 27
5.2 Variables . . . 27
5.2.1 Dependent variable . . . 27
5.2.2 Independent variables . . . 30
5.3 Descriptive statistics . . . 35
5.4 Concluding market share model for the Azorean case study . . . 36

6 Results **39**
6.1 Market share model . . . 39
6.1.1 Results of the market share model after liberalization including all airline-related independent variables . . . 40
6.1.2 Results of the market share model after liberalization without $freq_{day}$ as independent variable . . . 44
6.1.3 Results of the market share model after liberalization with monthly dummy variables . . . 46
6.1.4 Results of the market share model before liberalization including all airline-related independent variables . . . 49
6.1.5 Results of the market share model before liberalization without $freq_{day}$ as independent variable . . . 51
6.2 Impact . . . 53
6.2.1 Impact of ticket price on market share . . . 54
6.2.2 Impact of frequency on market share . . . 55
6.2.3 Impact of the independent variable $time\_points$ on market share . . . 56
6.2.4 Impact of the independent variable on_time on market share . . . 57
6.2.5 Impact of the monthly dummy variables on market share . . . 57
6.2.6 Impact of the constant on market share . . . 58
6.3 Deleted observations . . . 58
6.4 Fit of the model . . . 59
6.5 Summary of the results . . . 61

7 Sensitivity Analysis, Verification and Validation **63**
7.1 Sensitivity Analysis . . . 63
7.2 Verification . . . 64
7.3 Validation . . . 64
7.3.1 Results of the validation of the market LIS-OPO after liberalization . . . 65
7.3.2 Results of the validation of the market LIS-PDL-LIS after liberalization . . . 66
7.3.3 Results of the validation of the market OPO-PDL-OPO after liberalization . . . 67
7.4 Concluding remarks about testing the model . . . 67

8 Conclusions, Reflection, and Recommendations **69**
8.1 Conclusion . . . 69
8.2 Limitations and Reflection . . . 71
8.3 Recommendation . . . 72

A Correlation matricies **75**
A.1 Correlation matricies of the markets in the situation after liberalization 75
A.2 Correlation matricies of the markets in the situation before liberalization 76

B Results of the market share model for the situation after liberalization **79**
B.1 Results of the market share model with all airline-related variables 79
B.2 Results of the market share model without $freq_{day}$ as independent variable. 82
B.3 Results of the market share model adding monthly dummies 83

C Results of the market share model for the situation before liberalization **85**
C.1 Results of the market share model with all airline-related variables 85
C.2 Results of the market share model without $freq_{day}$ as independent variable. 86

Bibliography **89**

List of Figures

2.1 S-curve function [6] . . . 5
2.2 Nested logit structure of traveler choice Wei and Hansen [50] . . . 8

3.1 Visualisation of the development of the conceptual model . . . 14

5.1 The Azorean archipelago [5] . . . 26
5.2 Visualization of the calibration data . . . 28
5.3 S-shape of market share vs. η . . . 37

6.1 Relationship between raw data of ticket price variables and market share . . . 41

7.1 Actual market share vs. prospected market share in the combined market after liberalization for the market share model with all airline related variables . . . 66

A.1 Correlation matrix of the combined market after liberalization . . . 75
A.2 Correlation matrix of the route LIS-PDL-LIS after liberalization . . . 75
A.3 Correlation matrix of the route LIS-PDL after liberalization . . . 76
A.4 Correlation matrix of the route PDL-LIS after liberalization . . . 76
A.5 Correlation matrix of the route OPO-PDL-OPO after liberalization . . . 76
A.6 Correlation matrix of the route LIS-PDL-LIS before liberalization . . . 76
A.7 Correlation matrix of the route LIS-PDL before liberalization . . . 77
A.8 Correlation matrix of the route PDL-LIS before liberalization . . . 77

B.1 Results of the combined market for the the market share model with all airline-related variables after liberalization . . . 79
B.2 Results of the route LIS-PDL-LIS for the the market share model with all airline-related variables after liberalization . . . 80
B.3 Results of the route LIS-PDL for the the market share model with all airline-related variables after liberalization . . . 80
B.4 Results of the route PDL-LIS for the the market share model with all airline-related variables after liberalization . . . 80
B.5 Results of the route OPO-PDL-OPO for the the market share model with all airline-related variables after liberalization . . . 81
B.6 Results of the combined market for the the market share model without $freq_{day}$ after liberalization . . . 82
B.7 Results of the route LIS-PDL-LIS for the the market share model without $freq_{day}$ after liberalization . . . 82
B.8 Results of the route PDL-LIS for the the market share model without $freq_{day}$ after liberalization . . . 82

B.9 Results of the route OPO-PDL-OPO for the the market share model without $freq_{day}$ after liberalization . . . 83
B.10 Results of the route LIS-PDL-LIS for the the market share model with monthly dummies after liberalization . . . 83
B.11 Results of the route OPO-PDL-OPO for the the market share model with monthly dummies after liberalization . . . 83

C.1 Results of the route LIS-PDL-LIS for the the market share model with all airline-related variables before liberalization . . . 85
C.2 Results of the route LIS-PDL for the the market share model with all airline-related variables before liberalization . . . 86
C.3 Results of the route PDL-LIS for the the market share model with all airline-related variables before liberalization . . . 86
C.4 Results of the route LIS-PDL-LIS for the the market share model without $freq_{day}$ before liberalization . . . 86
C.5 Results of the route LIS-PDL for the the market share model without $freq_{day}$ before liberalization . . . 87
C.6 Results of the route PDL-LIS for the the market share model without $freq_{day}$ before liberalization . . . 87

List of Tables

2.1 Link function Olsson [31] 9

5.1 Public service obligations 26
5.2 Attribute selection 31
5.3 Descriptive analysis for the variables in the situation after liberalization 36
5.4 Descriptive analysis for the variables in the situation before liberalization 36

6.1 Results of the market share model after liberalization with all airline-related variables . . 40
6.2 Results of the market share model after liberalization without $freq_{day}$ 44
6.3 Results of the market share model after liberalization with monthly dummies 47
6.4 Results of the market share model before liberalization with all airline-related variables . 49
6.5 Results of the market share model before liberalization without $freq_{day}$ as independent variable 51
6.6 Impact of the variables in the market with all airline-related variables 53
6.7 Impact of the variables in the market with the airline-related variables disregarding $freq_{day}$ 53
6.8 Impact of the variables in the market with monthly dummy variables 54
6.9 Results of the market share model including monopoly observations 59
6.10 Results of the model fit calculation 60

7.1 Results of the sensitivity analysis 64
7.2 Results of the validation after liberalization 66

8.1 Significant independent variables 70

List of Acronyms

ETOPS Extended-range twin-engine operational performance standards
EU European Union
GLM Generalized linear model
HS Hub-and-spoke
IIA Independence of irrelevant alternatives
IWLS Iterative Weighted Least Squares
LCC Low-cost carrier
LIS Lisbon
OPO Porto
PDL Ponta Delgada
PSO Public Service Obligations
QSI Quality of service
TEN Trans-European high-speed rail network
TER Terceira
US United States of America

List of Symbols

Greek symbols

α	Significance level
β	Beta - coefficient matrix
χ^2	Pearson's chi-square
η	Linear predictor
μ	Random variable
ω	Weight function
∂	Partial derivative
ε	Error component

Symbols

D	Deviance
$freq$	Frequency
g	Link function
H	Hessian matrix
l	log-likelihood function
on_time	On-time performance
P	Probability
$price$	Ticket price
$time\_points$	Time points (number of flights in a preferred time window)
U	Utility function
V	Observed term
v	Variance
w	Weights
X	Independent variable matix
Y	Dependent variable matrix
z	Working response variable matrix

Subscripts

$\_2m$	two months prior
$\_day$	per day/day prior
$\_i$	alternative
$\_week$	per week

1 Introduction

In the early years of air transportation there were no international agreements that regulated the civil air traffic. Due to security reasons, there was a need for regulation when the popularity of air transportation increased. The basis of the regulation for international civil air transportation is the ’Chicago convention’ framework which was adopted in 1944.

In the following decades the air transportation market experienced a large increase. The resulting congestion at the airports, multiplicity of airfares, and many other effects resulted in the introduction of the ’Airline Deregulation Act´ in 1978 which liberalized the domestic US air transportation market Cohas et al. [11]. Another major step of deregulation was the ’open skies agreement´ between the US and Europe in 2008. This development resulted in less involvement from the government, but more competition between the airlines. Therefore, the airline industry experienced more pressure financially, which led to lower ticket prices and emerging low-cost carriers (LCC’s), but also many bankruptcies, for example PamAm in 1991. To escape the threat of a bankruptcy, numerous airlines have merged, like Air France and KLM in the Europe. [6]

In this research the Portuguese archipelago of the Azores is taken as a case study, because it is a recently liberalized low-demand market and collaboration with the Azorean airline SATA was possible. The EU commission introduced Public Service Obligations (PSOs) for air traffic in remote European areas like the Azorean archipelago. In those areas, air transport is vital for the infrastructure but often the demand is too low to operate profitable routes. The archipelago experiences low-demand due to its low population density and its remote location. If a route is PSO regulated, the EU offers a monopoly position on a specific low-demand route and will compensate the airline for the loss in revenue, if present. This is a contrary development comparing to the deregulation in the air traffic market but it is necessary for the inhabitants of these remote locations to have flights regularly.

However, it also happens that routes in a PSO regulated market start to gain passenger traffic such that they are continuously profitable. This happened on the routes between Ponta Delgada and the Portuguese mainland. In these situations, the EU and the government can choose to liberalize the market which was dictated by the PSOs for many years. If the route is promising , the airline in the monopoly market has to expect new competitors to enter the market, which in the case of the Azores are: Ryanair and easyjet. If a low-cost carrier participates on a route, it can offer highly competitive ticket fares for the passengers.

Liberalization results in the fact that airlines that were operating in a monopoly market before, need to adjust their operation schedule and pricing for the liberalized situation. However, the airline, in this case SATA, often has little to no experience with aggressive competition. Especially if a low-cost carrier enters the market, the legacy carrier is unable to compete with the competitor on price level and considers changing the service offered. In the new situation it is unclear which service attributes are important when it comes to passenger choice behavior. This research can help airlines in a newly liberalized market to make right strategic and tactical decisions to position their airline appropriately in the market. Next to the application to industry, the research also has aims to expand the body of knowledge by developing a market share model with a methodology that was not applied earlier in this aeronautical context.

The research goal of the MSc thesis is to develop a framework that is able to **determine the short-term impact of liberalization on a newly liberalized, low-demand market, that can be applied to the Azorean archipelago as a case study**. Therefore, a market share model is developed that is able to process airline characteristics (attributes) like ticket price, frequency, service etc. as input variables and predict the market share for the liberalized situation. The model gives an indication about the impact of each attribute on the airline market share. Also, the influence of each attribute is compared for the situation before and after the liberalization. Only the short-term impact of liberalization could be investigated right after the liberalization since it is such a recent example of liberalization.

The thesis report consists of 8 chapters and is focused on how to develop the market share model. After this first introductory chapter, Chapter 2 will provide an overview of the literature on suitable methodologies and the liberalization context. From there, a research question was established that lays the foundation of the research. Together with the conceptual framework and the relevance of the thesis it forms the research methodology which can be found in Chapter 3. In Chapter 4 the market share model is described in full detail. This model is then applied to the case of the Azorean archipelago which is dealt with in Chapter 5. There, all the steps are described to built the model and to generate the results that are presented in Chapter 6. The model is applied to different routes between the Azores and the Portuguese mainland before and after the liberalization which can be seen in the structure of the chapter. In Chapter 7 the sensitivity analysis, verification, and validation of the model is explained to ensure that the model is correct. In the last chapter, conclusions are drawn and recommendations are given about how to improve the model.

2

Literature study

Before starting to develop the market share mode, it is necessary to investigate the situation and which methodologies can be used to create this model. This literature study has been performed to gain a general overview of the literature dealing with different methodologies to construct market share models. Furthermore papers are investigated that deal with the newly liberalized situation and the influence of airline attributes on market share.

First, the liberalized situation is analyzed and the influence of the low-cost carriers is investigated in Section 2.1. Then, in Section 2.2, different methodologies are shown that have been used to construct market share or passenger choice models. In the end, the results are summarized and conclusion can be drawn in Section sec:results.

2.1. Liberalized situation and influence of low-cost carriers

Studies that investigated the impact of liberalization on air traffic show that due to liberalization the passenger traffic will increase but the ticket price and therefore yield will decrease. Adler et al. [3] conclude from the analysis that the competition on a newly liberalized market leads to an increase in frequency from the legacy carriers. Furthermore, they found out that the ticket prices of all airlines decrease compared to the case before liberalization. Also, it can be seen that if the airline has a choice (which is usually not the case for low-cost carriers), they choose to fly smaller aircraft. Since the low-cost carriers can offer highly competitive prices, the legacy carriers have to improve quality to sustain their position. All passengers traveling in the market will benefit from welfare gains through increase in frequency and hence service quality, and the reduction of ticket fare.

This results from the fact that competition increases so there is more pressure on the pricing and service quality of the airline Dresner et al. [14]. Based on an empirical analysis of the deregulated EU air traffic network Janić [22] concluded that due to the increase of frequency on certain routes the schedule delay is reduced. The reduced ticket price and schedule delay are major gains for the passengers which are caused by liberalization.

Next to stimulation of the already mentioned traffic growth, Fu et al. [17] stated that also economic growth is triggered by liberalization due to employment opportunities, trade promotion and better transport and logistic services. O'Connell and Williams [30] concluded from analysis that legacy carriers are not able to match the fares of the low-cost carrier due to, amongst other things, inefficient operating practices. To cut some costs legacy airlines tend to optimize their networks after liberalizations according to Fu et al. [17]. This means that the legacy carriers try to move the network more into the direction of a 'hub-and-spoke' (HS) network or they try to find the most optimal HS network.

Legacy carriers are not expected to enter a local market of another legacy carrier since they foresee a loss in profit due to retaliation and network effects [35]. Low-cost carriers do not have the network effects because they perform their flights in a 'point-to-point' manner which does not include transfer passengers. The paper also addresses further properties of low-cost carriers like the single aircraft strategy, low turn-around times and direct sales of the tickets which help this airline type to keep the costs low so that they can offer highly competitive prices. Also, low-cost carriers use secondary airports which passengers are willing to connect through to accept no frills according to O'Connell and Williams [30].

2.2. Methodologies and impact of airline attributes on market share

In this section different methodologies are presented that have potential to be used for the market share model. The papers that apply the methodologies investigate the impact of airline attributes on market share. Therefore, these results can also be found in this section.

The first methodology, that is examined, is the quality of service index which is found in subsection 2.2.1. In the following subsection 2.2.2 the logit model in explained. The subsection is divided in three parts for the three different types of logit models namely the binary logit model in 2.2.3, multinomial logit model in 2.2.4 and the nested logit model in 2.2.5. Additionally in subsection 2.4, game theory is described in more detail. Finally the aggregate-level Markovian type model and fuzzy logic is examined on its usability as a methodology for the market share model in subsection 2.5.

2.2.1. Quality of service index

The quality of service index (QSI) was firstly introduced by the US government in 1957 and is now widely used in the airline industry. The model can calculate the market share of an airline in comparison to other airlines on the same route. The passengers compare the characteristics of the airlines, which are considered to be the quality of service, and come to the conclusion which airline to choose. The quality of service is a function of independent variables and their corresponding preference weights. The function can be either linear or multiplicative as can be seen in equation 2.1[4]. Other forms of that function are also possible. The attributes of the airlines are represented by independent variables which are denoted with X_i whereas the preference weights are indicated with β_i.

$$QSI_i = \sum_i \beta_i X_i \quad or \quad QSI_i = \prod_i \beta_i X_i \tag{2.1}$$

$$S_i = \frac{QSI_i}{\sum_{j \in J} QSI_j} \tag{2.2}$$

The characteristics of the itinerary can be objective like frequency, number of stops or seats allocated but can also be subjective since some attributes are perceived differently by different passenger types. Subjective attributes can be ticket price, airline brand or type, availability of a frequent-flyer program etc.. The preference weights can be defined with statistical methods (calibrated with past data) or analyst intuition. The calculated value for the quality of service can then be used to calculate the passenger share S_i of airline i with the help of equation 2.2. The numerator consists of the QSI of the airline under investigation while in the denominator, the QSI of all the airlines in the market are summed up. [4]

If considering the relationship between frequency share (share of the total flights) and market share (share of the market also influenced by other airline attributes) with all other attributes equal, the typical S-curve share will be present. This means that if the frequency share is 50% the market share is expected to be 50% too. If the frequency share is higher the gains will increase rapidly whereas after around 75% the returns are diminishing as can be seen in figure 2.1. This relationship is often used as a rule of thumb in the airline industry. [6]

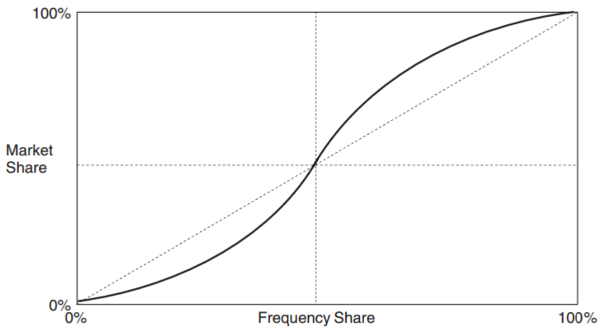

Figure 2.1: S-curve function [6]

The reason why this model is widely used in the airline industry is because it is simple. However there are also disadvantages of the model. One problem is the determination of the preference weights. These are usually determined independently of other preference weights which results in the fact that correlation between different airline attributes is not captured in the model. Furthermore, it might be problematic if the preference weights are based on analyst intuition. Additionally, the model does not take into account that there exist competition between the airlines so that attributes can be adjusted according to changes the opponent made. [4]

2.2.2. Logit model

Berkson [8] was the first scientist to present the logit model as an economic methodology. It is based on discrete choice theory and predicts choices of decision-makers (individuals, groups, governments etc.) between different alternatives. The decision-maker bases its choice on the attributes of the alternatives, which are characteristics of the alternatives, according to the terminology of Domencich and McFadden [13]. The logit model uses random utility theory to pick the most probable alternative to be chosen, this is described in great detail in Ben-Akiva and Lerman [7]. According to this model, the chosen alternative is the one that maximizes the utility. The utility can be calculated using Equation 2.3.[18]

$$U_i = V_i + \varepsilon_i \tag{2.3}$$

$$V_i = \beta' X_i \tag{2.4}$$

$$P_i = \frac{expV_i}{\sum_{i=1}^{n} expV_i} \tag{2.5}$$

The utility function U_i for different alternatives i is split into two components. The first part is the observed term V_i which is further defined in equation 2.4. The equation includes unknown parameters β' and an array of a set of attributes (X_i for the different alternatives i). The ε represents the error component of the equation. It is a random variable that takes into account everything that is unknown but may have influence on the outcome of the utility function. The unobserved error component is assumed to be independent and identically Gumbel distributed. [18]

The probability P_i that an alternative i is chosen can be calculated using equation 2.5. It represents the theoretical market share of the different airlines choices that needs to be determined. The function also represents the S-shape relationship which is typical for the airline market share . Even though the shape is the same, the formula is different from the QSI-model. [18]

There are three different types of logit models: binary, multinomial and nested. The binary logit model is the least complex of the three and can be used if there are two possible alternatives. The multinomial logit model predicts the choice between multiple options which is developed from the binary logit model. The nested logit model is an appropriate model choice if the outcomes are interrelated.

2.2.3. Binary logit model

As described above, the binary logit model is an appropriate choice if there are two alternatives to choose from. These alternatives represent for example success and failure, yes and no or other dichotomous data. In the context of airline choice models a binary logit model was presented in the paper of Suzuki et al. [44]. It argues that an airline market share model can be simulated using a binary logit model since the market share is a number between 0 and 1. The proposed model addresses the relationship between airline market share and customer service quality. It is found that the attractiveness of the air carrier is influenced by price and the service quality of the airline which is calculated by a weighted average function.

Even though Suzuki et al. [44] proposed the binary logit method, the author found out that a linear regression model provides a better fit for the relationship between market share and airline service quality. When looking at the market share values, a stronger negative effect can be seen when the service quality is below the reference point than the other way around. This phenomenon is called negativity bias or loss aversion and is proven to be present for service quality.[44]

2.2.4. Multinomial logit model

The multinomial logit model can be used to solve passenger choice models with three or more alternative outcomes. The methodology is a development from the binary logit model and was greatly influenced by several works of Daniel McFadden ([25], [26]). The latter paper deals with the biggest downside of the multinomial logit model namely the assumption of independence of irrelevant alternatives (IIA).

Proussaloglou and Koppelman [37] use the multinomial logit model to examine the influential factors on air carrier demand and air carrier choice. The paper is based on a passenger survey about their choice behaviors. The utility function of the model is dependent on the following attribute groups of the airline alternatives: carrier market presence, level of service, quality of service and pricing. The utility function and the market share are calculated in the same way as described in section 2.2.2. The conclusion shows that all attribute groups have significant impact on the carrier choice. The strongest influence has the participation of a frequent-flyer membership which stimulates carrier loyalty. Next to that, schedule convenience, low ticket fares (especially for leisure passengers) and on-time performance (especially for business passengers) were determined to strongly influence passenger airline choice. However, only a balanced mix of those attributes result in maximum utility for passengers.

Román et al. [39] analyze the mobility of the Spanish and Portuguese archipelagos: Canaries, Madeira and Azores. The authors base their model on a passenger survey for the routes Gran Canary - Madrid, Funchal - Lisbon and Ponta Delgada - Lisbon in 2005. This data is used to make a multinomial logit model to find out which service attributes are significant when making airline choices. It was concluded that next to airfare, reliability and comfort as well as food are valued attributes. The reliability aspect combines on-time performance with reimbursements whereas comfort is defined as in-flight seat space and food. A new perspective that can be found in Román et al. [39] is that the value for the attributes is converted to monetary units. Additionally, the authors state that there is a significant difference between economy and business passengers which results in the fact that they should be addressed differently in the utility function of the logit model.

Levy and Panou [24] investigates another remote island location by taking the Greek island Chios as a case study. The author concludes from the analysis of the passenger survey that passenger choice is dependent on price for the trip, travel time and frequency which is in line with other findings presented in this section. The focus on this study lies on the hypothesis that inhabitants of the island will cancel or reschedule their trip to the Greek mainland if their preferred alternative, in terms of ticket price, schedule convenience and trip time is not available. This hypothesis was found to be true and additionally the results show that a higher ticket price cannot be compensated by shorter travel time.

In Carrier [10] the choice of itinerary and ticket price is analyzed with the use of booking data extracted from the airline reservation system Amadeus. The author found a way to determine passenger type and choice without finding the preference data via, for example, a passenger survey. A stated preference survey produces data that might have the risk to bias the representation of the passengers. However, the problem with this approach is that only the booked alternative is recorded in the booking data. This means that the booking data does not capture between which alternatives the passenger was choosing.

2.2.5. Nested logit model

An extension of the binary and the multinomial logit model is the nested logit model. The nested logit model does not have the downside the multinomial logit model has, which is that he attributes need to be independent from irrelevant alternatives according to McFadden et al. [26]. This means that the nested methodology is the appropriate choice if the alternatives are interrelated. It clusters options that are interrelated into different nests which means that the choices have to be made in several steps.

Wei and Hansen [50] tries to find a relationship between aircraft size and seat availability as well as airlines' demand and market share in a duopoly market. The author used a nested logit model with 'No air travel' in one nest and the choice between legacy and low-cost carrier in the other nest (Figure 2.2). The paper addresses the problem that if the passenger demand increases the airlines need to decide on a strategy to either increase frequency or aircraft size. Increasing the frequency results in the fact that airports tend to get congested.

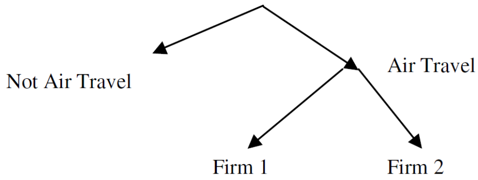

Figure 2.2: Nested logit structure of traveler choice Wei and Hansen [50]

Wei and Hansen [50] came to the conclusion that an airline achieves a higher market share more from increasing frequency than from increasing aircraft size. A higher flight frequency attracts more passengers but there is no significant difference in attractiveness for the passenger between a small aircraft with higher percentage of seats available and a big aircraft with a low percentage of seats available given that the net value of seats is the same.

Adler et al. [3] use the same methodology as Wei and Hansen [50] but distinguishes between hub-spoke or low-cost alternatives in the 'air travel' nest and between traveling by rail, road or not traveling at all in the 'not air travel' nest. The model was created to capture the effect of international air transport liberalization, regional open skies policies, domestic inter-modal competition and airport slot allocation on air traffic in the Northeast Asian market.

2.3. Generalized linear model

Generalized linear model (GLM) is a theory that generalized the general linear model and was firstly introduced by Nelder and Wedderburn [29]. While the general linear model assumes that the response variable is independently normally distributed, the GLM allows the response variable to be normal, Poisson, gamma or binomial distributed. In equation 2.6 the general form of the GLM can be found.

The linear predictor part of the model in the matrix form is $X\beta$ which is defined as η in equation 2.7. $g(\mu)$ represents the link function which is another difference to the general linear model. The link function connects the expected values of the dependent variable Y to the independent variables X. The most common forms of the link function can be found in table 2.1.Olsson [31]

$$Y = \beta X + \varepsilon \tag{2.6}$$

$$g(\mu) = \eta = \beta X \tag{2.7}$$

To solve the GLM equation maximum likelihood estimation is applied. This method finds the regression coefficients that maximize the log likelihood through an iterative process. The inputs of the methodology are values for the dependent variable matrix Y and the independent variables matrix X. The output is an estimation of the β-coefficients which can be compared to each other. The β-coefficient is linked to its attribute so only in combination conclusions can be made about the importance of the respective attribute. If only the dataset is changed and the model is the same, the results of the methodology can say something about how the attribute importance changes due to change of the situation.

Table 2.1: Link function Olsson [31]

Distribution	Family	Link function	Inverse
Normal	identity	μ	η
Inverse Gaussian	Inverse squared	$-\frac{1}{\mu^2}$	$\frac{1}{\sqrt{-\eta}}$
Binomial	logit	$ln(\frac{\mu}{1-\mu})$	$\frac{exp(\eta)}{1+exp(\eta)}$
Poisson	log	$ln(\mu)$	$exp(\eta)$
Gamma	inverse	$-\frac{1}{\mu}$	$-\frac{1}{\eta}$

One of the major advantages of the GLM is that it is able to work with proportions as dependent response variable. This characteristic was introduced by Papke and Wooldridge [32] who applied the GLM to analyze the participation rate of a pension plan. The participation is given as a percentage which is bound between between 0-1, similar to the market share. Furthermore, many statistical software packages exist that have the GLM implemented. This results in the fact that the GLM methodology is easy to use and it may not be necessary to program the model which will make a difference regarding time.

2.4. Game theory

Game theory was firstly introduced by Von Neumann [47] and greatly influenced by John Forbes Nash with papers like Nash [27] and Nash [28] in which he developed the Nash-equilibrium. The theory represents the strategic interaction between different parties like individuals or organizations and is often used in cases of decision making. Game theory can be applied to the transportation market when passengers have to choose between different alternatives.

The participants in the game are called *players* that have different *strategies sets* available to achieve gain or loss in the *payoff function*. There are different game theory models available which differ in the *order* of a decision from an player (e.g. leader-follower game), in the independence of the decision choices from the other players (e.g. non-cooperative game) and the influence of one decision on the other (e.g. Nash equilibrium) [52].

In Wei and Hansen [51] game theory is linked to airline competition in such a way that airlines (*players*) choose their attributes (*strategies set*) and then the costs and market share (*payoff function*) of each airline is calculated. The paper takes the cost function of Wei and Hansen [49] and the market share model of Wei and Hansen [50] and applies it to three game-theoretic models. These models analyze airline choices in a duopoly market. The authors focus on aircraft size and service frequency.

The main finding of this paper is, that in every solution for the short-haul market, the airlines choose the smallest available aircraft and therefore increase service frequency to comply with rising travel demand. This result is in line with the conclusion in Wei and Hansen [50] that states that if airlines increase their service frequency they achieve higher market share increase than from increasing aircraft size. But due to airport capacity there might be a situation where the airlines have to be forced to use bigger aircraft.

In the paper of Hsu et al. [21] the different alternatives under investigation are two railway companies, car, bus and aircraft. These transportation modes are competing on a mainly parallel running track on the north-south coast of Taiwan. The passengers choose the transport mode and the according route by calculating the total travel cost (*payoff function*) that consist of ticket price, travel and transfer time cost, and access and egress cost.

After simulating the Nash equilibrium Hsu et al. [21] draw the conclusion that if passenger distribution along the track is reduced, the competition between the two railway companies decreases too. This results in the fact that the ticket prices will be raised. Another important parameter of the equations is the travel time cost. Business passengers tend to have a higher time value thus a higher travel time cost. These passengers are inclined to choose the high-speed rail. If the demand for high-speed company increases due to an increase in travel time costs, so does the computed ticket price for this company. The low-speed railway company on the other hand needs to reduce prices to keep its market share.

The paper of Adler et al. [2] uses game theory to analyze the competition between hub-and-spoke legacy carriers, low-cost airlines and high-speed railways taking into account the railway expansion projects ”Trans-European high-speed rail network” (TEN) up until the year 2020. The market share is computed for each transport provider separately and follows the nested-logit method as explained in Section 2.2.5. One nest consists of air travel and the other nest represents the options to use the high-speed train or not travel at all. The air travel nest is divided in the legacy carrier and low-cost airline options similarly to the model used in Adler et al. [3].

The model develops passenger transport market equilibria using game theory. It can be concluded that the rail operations are highly dependent on the access costs, the operator has to pay for infrastructure use. The profits for the railway company are significantly higher when charging low access charge and taking the TEN expansions into account. The same counts for low airport charges. Furthermore, the railway companies can increase their ticket price almost to the level of the low-cost carriers if considering the trip time from city center to city center because if landing on a secondary airport with a LCC it takes time and money to reach the city center.

2.5. Other methods

There are also other methods that are suitable to relate market share to different airline characteristics. An example is the aggregate-level Markovian type model that is discussed in Suzuki [43]. This model has different states and investigates the probability of switching to another state. The author investigates the relationship between on-time performance and airline market share. The model differentiates between passengers that have experienced flight delay (more than 15 minutes) in the last month and those who did not, which are the two states of the Markovian model. Then it is analyzed if the passengers that have experienced flight delays are more likely to switch to other airlines than the other passenger group.

It is shown that passengers are likely to be loss averse, which means that they show a stronger reaction to a negative outcome than to a positive, which was also found in Suzuki et al. [44]. Passengers expect airlines to be on-time, but when the airline is not on-time, passengers show an increasing switching rate. If a passenger did not experience delay the months before, it will tend to choose the airline which is dominant at the desired airport.

Another method present in literature is the fuzzy logic introduced by Hsu and Wen [20] to airline flight frequency determination linked to competition. The model calculates the optimal flight frequency and basic airfare for a particular origin-destination-pair of China Airlines in Taiwan in a competitive environment. The airline competitive interaction model is based on fuzzy logic which means that the attributes (in this case market share, changes in flight frequency and changes in airfare) are not the exact numerical value but a fuzzy set (like ’small’, ’medium’ and ’large’). The authors conclude that the model provides a reasonable but not optimal fit for the case study of China Airlines.

2.6. Methodology discussion and selection

In this section the different methods are analyzed based on the literature review of section 2. The end of the review is discussed first. The aggregate-level Markovial type model was used in Suzuki [43]. The model suffers from the assumptions made which make it hard to apply it to another, complex context.

Fuzzy logic got linked to flight frequency determination by Hsu and Wen [20] but it also provides a market share model. The authors conclude that the model provides only a reasonable fit in the case study. One reason for that might lie in the model itself because it clusters outcomes in (here 3) fuzzy sets which leads to loss of accuracy.

Furthermore, there were different papers found on airline market share and game theory. They show the applicability of game theory in the context of passenger airline choice. The model can be used for comparing attributes to another. However, game theory becomes complex with many players involved.

There are different types of logit models which are explained in detail in Section 2.2.2. The binary logit model was concluded to be not useful for the intended application. The papers on multinomial logit model use passengers surveys or actual booking data as data input. However, it does not seem feasible to conduct a passenger survey or receive booking data during this project. The nested logit model clusters similar alternatives into nests so that it can be used for interrelated attributes. This methodology does seem to be appropriate for the market share model that need to be constructed.

The following methodology is the generalized linear model. It makes use of the logit model and there are statistical programs available for generalized linear models which makes implementation of this theory easier. Furthermore, it seems applicable for the proportional data which is expected to be available for the project.

Finally, the quality of service index is described as a methodology to construct the market share model. The upside of this model is its simplicity. The major downside lies in the determination of the coefficients which might be inaccurate if selected by analyst intuition but also, no interaction between coefficients can be measured because of the lack of real data.

The methodology that seems to be most promising is the generalized linear model. It makes use of the logit model in the link function but preferred because it is able to process proportional data as an input. No literature could be found which addressed low-demand markets, liberalization, market share modeling and generalized linear model simultaneously which will be the focus of this MSc thesis in combination with a case study on the Azorean archipelago.

3
Research Methodology

A structured approach for a thesis project is needed to achieve a successful process. This research methodology chapter gives an overview of the steps that have been taken before constructing the market share model. The driving factor of the thesis is the research question which states the questions that need to be answered before the project is considered to be finished. Formulating the research question is one of the first steps to take for a successful thesis. The research question for this MSc thesis can be found in Section 3.1. After that, the conceptual model is shown. Section 3.2 gives an overview of the steps that needs to be taken to create the market share model. Finally, in Section 3.3, the relevance of this thesis is discussed.

3.1. Research question and objectives

The research question is the guideline for the thesis project. At the end of this project, the following research questions should be answered using the developed market share model. To structure the research, the main question is divided into four sub-questions. As can be seen below, the sub-questions are also split up.

What is the short-term impact of liberalization on airline market share in a newly liberalized low-demand market?

1. Which attributes can be identified that have significant impact on the market share in the described market?
2. What is the influence of the significant airline attributes on market share?
3. How does the influence of the airline attributes change due to the liberalization?
4. What is the probability that passengers choose for a specific airline in the described market in the future?

From the research question, the research objective can be drawn. The research objective is to develop a framework that is able to quantify market share for the different airlines involved in liberalization, by making a prediction about the short-term market developments. For this, airline-specific-attributes such as ticket price, frequency, and service are taken into account. In the end, it should be possible to compare the effects of the different attributes on the market share, and assess the impact of each variable before and after liberalization.

3.2. Conceptual model

Based on the research question, the conceptual model is created. A visualization can be found in Figure 3.1. It shows that the first step towards a successful market share model is the determination of the appropriate methodology. This selection takes into account that airline attributes are suppose to be the input data for the model and that in the end, the model should be capable to compute market shares. In order to find a feasible methodology, a broad literature review is carried out. In this literature study, the GLM is selected as the methodology.

In the next step, the variables are selected for the GLM. The selection is based on the adaptability to the case study. The market share model is applied to the Azorean archipelago, which is assumed to be representative for a newly liberalized, low-demand market. The model should be able to implement airline attributes as variables. The variable selection is also based on literature, but extended by own ideas.

Furthermore, data needs to be collected for the variables. For this project, collaboration is possible with the Azorean airline SATA. SATA is able to provide the majority of the data. Variables can be selected if data is present. Therefore, a loop is created in Figure 3.1 that indicates that the variable selection need to be updated after it is clear which data is available. The data is analyzed, sorted and put into a processable data structure.

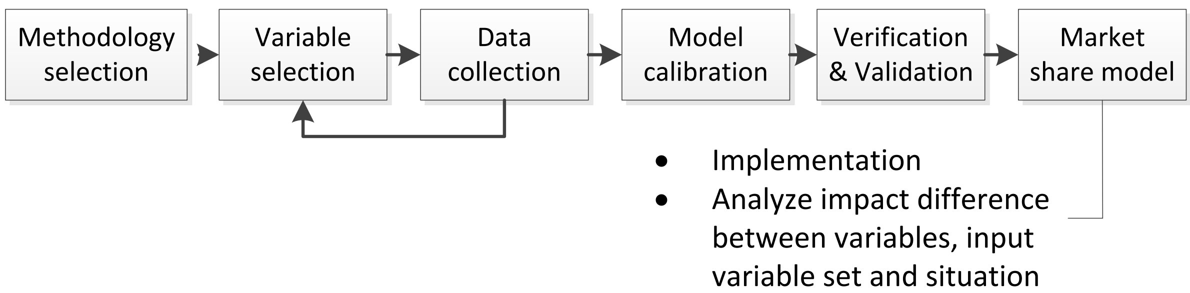

Figure 3.1: Visualisation of the development of the conceptual model

The model is calibrated with the data that is collected. This is the input for the GLM calculation. The GLM determines coefficients for every independent variable. With that, a preliminary market share model can be created. Results are produced which need to be verified and validated. To do that, the prospected market share is compared to the validation data. Additionally, it needs to be checked if the model actually solves the problem statement that was defined for this project.

It is possible to draw conclusions about how the market has changed due to the liberalization with the calculated coefficients. Also, the change in impact between the variables can be determined, for different input variable sets and for the situation before and after liberalization. Furthermore, when the theory is implemented in the framework it is possible to make predictions about the short-term market developments, which is the research objective of the project.

3.3. Relevance

This thesis creates a market share model that can to be applied to a newly liberalized low-demand market. Several papers were found that link liberalization to a market share model or look at the influence of attributes on each other as can be seen in Section 2. However, the GLM with logit link, with which the market share model is constructed in this thesis, has not been applied on a newly liberalized low-demand market in transportation research before. The GLM with logit link was applied to employee participation rates in Papke and Wooldridge [32]. Also, the logit model is applied to several airline passenger choice models like for example Adler et al. [3]. Furthermore, the impact of liberalization is addressed before in Adler [1]. But a market share model that investigates the impact of liberalization has never been built with the GLM with logit link. This is a novelty to the body of knowledge.

The model is created to serve the needs of the Azorean archipelago which is taken as a case study. But when the model proves to give satisfying results it could also be applied to other liberalization processes in other low-demand markets. European examples for that may be the case of small island groups in Norway or Greece.

From an industry point of view the outcomes of this research are also considered to be valuable. The routes to the Portuguese mainland are liberalized and there is competition from two low-cost carriers that can offer highly competitive prices. The Azorean airline SATA will have a hard time to stay profitable in the market. Especially because the airline is forced to provide connecting flights to all passengers entering the Azores for free. More about that can be found in Section 8.2. The existence of SATA is vital to the area because it is the only airline that operates flights between the nine different Azorean islands. Air transport is indispensable for the Azores in terms of mobility and the supply of the inhabitants.

Furthermore, the LCC Ryanair announced interest in flights to another Azorean island, Terceira. The liberalization of the public service obligations also included flights between Terceira and Lisbon. There were ongoing negotiations, but in September 2015 the government denied Ryanair to access the airport of Terceira. However, Ryanair still has interest on the route so the research might not only be applied to the flights from Ponta Delgada to the mainland of Portugal, but also to flights to and from Terceira if the situation changes. If the important airline attributes in the market are known beforehand, SATA can be better prepared for the liberalized situation.

The framework can give results about which attributes should be adjusted and which attributes only have minor influences on market share. Therefore, resources can be invested in the attributes that matter and the company is able to adjust their schedule according to the prediction of the model. Also, the fleet and personnel management decisions can be made accordingly.

4

Market share model

The Generalized Linear Model (GLM) is chosen in the previous section to be the most suitable methodology to build the market share model for the newly liberalized low-demand market. The characteristics of the GLM are briefly described in Section 2.3 but are elaborated in more detail in the following. For this thesis, the statistics software STATA® is used which is able to calculate the results of the GLM after defining the input variables. This section focuses on the theory of the GLM method and shows how to calculate the β-coefficients that indicate the importance of the corresponding airline attribute on market share.

First of all the structure of the market share model is explained in Section 4.1. In Section 4.2 the quasi-maximum likelihood estimation is evaluated. Furthermore, the model uses robust standard errors that are described in Section 4.3. In the following Section 4.4, the theory behind the p-value statistics is explained. Then, it is shown how the correlation of the variables can be determined in Section 4.5. The model needs to be tested on the goodness of fit. The different approaches can be found in Section 6.4. To draw conclusions it is necessary to determine the impact on market share as can be seen in Section 4.7. In Section 4.8, concluding remarks are given.

4.1. Structure of the market share model

The GLM is chosen to construct the market share model. In Equation 2.6, the general form of the GLM can be found and it consists of the dependent variable Y which is also known as the dependent variable, the independent variables X_i that describe the airline attributes, the β-coefficients and the error term ε.

Typical for the GLM are 3 components which are namely: the random component, the linear predictor and the link function. The random component describes the distribution of the dependent variable. Common distribution for the GLM are the Binomial, Poisson or Gamma distribution.

The next component is the linear predictor. This component describes the form under which the independent variables and the β-coefficients are ordered. It is denoted with the symbol η in Equation 2.7 and describes that the right hand side of the equation is of the form $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ etc..

The linear predictor goes hand in hand with the link function which is the third component of the GLM. It links the mean of the observations (μ) to the linear predictor η. The link function is needed because when extending the general linear model to the generalized linear model, the random component is not bound to the Normal distribution but can take over different distributions. Therefore, the range of the distribution changes. The link function adjust the mean of the observations in such a way that is properly links linear predictor [12]. The most common link functions can also be in 2.1.

In this thesis the model is built with the random component being fraction since it represents market share. Papke and Wooldridge [32] first mentioned the possibility of using fractional response data in the GLM. With the help of robustness, described in more detail in Section 4.3, the fractional dependent variable can be approximated to a Bernoulli distribution which is a specific (one step) case of the binomial distribution.

When using binomial distribution for the random component with a logit link function, the GLM can be used to create the market share model. Equation 4.1 shows the link function for the market share model. The link function is invertible and creates the mean function that can be found in Equation 4.2.

$$g(\mu) = \beta X = \eta = ln\left(\frac{\mu}{1-\mu}\right) \tag{4.1}$$

$$marketshare = \mu = \frac{exp(\eta)}{1 + exp(\eta)} \tag{4.2}$$

To answer the research questions that were defined in the previous, the β-coefficients need to be calculated. With the β-coefficients, the impact of the corresponding independent variable on market share can be determined. The independent variable X_i describes the airline attributes. This variable is explained in more detail in Section 5.2.1 but for the moment it is important to point out that the variable represents airline characteristics like ticket price or frequency. The data for this variable needs to be collected but in the end the values for the variables are known. Following straight from the fact that the β-coefficients need to be determined, the random component μ is unknown. [42]

4.2. Quasi-maximum likelihood estimation of a fractional response variable

Market shares are fractions that indicate the proportion of passengers that were transported in the market. When only one airline transports passengers then a monopoly situation is present and the market share is 1 whereas the competitor has a market share of 0 for the same time period. In this case a market share model is developed where a fraction (market share) is taken as the dependent variable.

Usually, in the case of the GLM, the β-coefficients are calculated by applying Iterative Weighted Least Squares (IWLS) estimation. It is used to find the maximum likelihood of least squares of the distances between the estimator and the sample. However, when the dependent variable is a fraction, this methodology is not applicable any more. To use fractional response data in the GLM, Papke and Wooldridge [32] developed a quasi-likelihood method (also known as pseudo maximum likelihood method) which will be presented here. To be more specific, the methodology that is used is called Bernoulli log-likelihood estimation that is shown in Equation 4.3.

$$l_i(\beta) = y_i log[G(x_i\beta)] + (1 - y_i) log[1 - G(x_i\beta)] \tag{4.3}$$

This Bernoulli log-likelihood equation is maximized to find the best suitable values for the β-coefficients.

This is done by taking the derivative of Equation 4.3 and setting is equal to zero. The β-coefficients that satisfy this formula are the ones that describe a maximum point in the function. The coefficient is evaluated in an interative process.

The function $G(\cdot)$ can take over various distributions. However, in this thesis the logit model is used. This means that the function $G(\cdot)$ is defined as shown in Equation 4.4 which is the same as Equation 4.2 derived in Section 4.1. It is important to note that the function satisfies the boundaries $0 < G(x_i\beta) < 1$ for all $x_i\beta \in \mathbb{R}$. This is a necessary assumption that has been made to develop the quasi-maximum likelihood estimation of a fractional response variable.

$$G(x_i\beta) = \frac{exp(x_i\beta)}{1 + exp(x_i\beta)} \tag{4.4}$$

This approach for estimating the β-coefficients is consistent and asymptotically normal for any kind of distribution the dependent variable can take over. This makes it possible to use the quasi-maximum likelihood estimation for fractional response data. The most important thing to notice in this section is that the true distribution of the entire model is not needed to be known to obtain consistent parameter estimates. [42]

4.3. Robust Standard Error

The standard error indicated how much the observations are spread from the mean. The higher the standard error, the more the observations deviate from the mean. Therefore, the standard error is needed to understand the observations are there given as input for the market share model. Additionally, the standard error is needed to calculate the p-value which will be explained in detail in Section 4.4.

To use the GLM, the independent variables and the error are assumed to be independent and identically distributed.Olsson [31] However, when creating the market share model, the independent variables are not necessarily identically distributed. This results in the fact, that the standard error of the fractional logit model estimated with the GLM is not valid any more. This problem can be solved by using the "Huber Sandwich Estimator" which is another name for the method of calculating the robust variance which is presented in the following [40]. Actually, it is not the robust variance that is of interest, but the robust standard error is needed for the GLM. However, the robust standard error can easily be determined by taking the square root of the robust variance [48]

For the case of a fractional response variable, the estimated variance ($\hat{V}(\hat{\beta})$) for an estimate of β (denoted as $\hat{\beta}$) of the GLM can be calculated with Equation 4.5. In Equation 4.5, H stands for the Hessian matrix defined in Equation 4.6, n stands for the number of observations and $l_i(\beta)$ for the Bernoulli log-likelihood equation defined in Equation 4.3.

$$\hat{V}(\hat{\beta}) = -H^{-1}\left\{\frac{n}{n-1}\sum_{i=1}^{n}\left(\frac{\partial l_i(\beta)}{\partial\beta}\right)'\left(\frac{\partial l_i(\beta)}{\partial\beta}\right)\right\}(-H^{-1}) \tag{4.5}$$

The parameter that is still missing to calculate the estimated variance is the Hessian matrix that is defined in Equation 4.6. The Hessian is the matrix of second derivatives which means in this case that it is the second derivate of the likelihood equation which is the Bernoulli log-likelihood equation when robust estimation for fractional response data is applied.

$$H = \frac{\partial^2 l_j(\beta)}{\partial^2 \beta} \tag{4.6}$$

All parameters of the estimated variance calculation are now known. If the square root is taken from the estimated variance, the robust standard error for the market share model can be evaluated. [40]

4.4. Wald test and p-value

The Wald test is used to determine the z-scores of the proposed model. It is used to test a set of coefficients on the null hypothesis which is defined as $H_0 : \beta = \beta_0$. That means that the test shows if a coefficient of the set, differs from 0 or not. So, ultimately the test is used to find out whether or not the coefficient has influence on the dependent variable; i.e. whether it is significant. The result of the Wald test is the squared z-score.

The calculation of the z-score can be performed with Equation 4.7. In the equation, $\hat{\beta}$ represents the estimated value for the β-coefficient and β_0 is the β-coefficient of the null hypothesis which is equal to zero. Furthermore, $\hat{V}(\hat{\beta}$ represents the estimate robust variance of the model. The square root of the estimated robust variance equals to the estimated robust standard error which was previously explained in Section 4.3.

$$z = \frac{(\hat{\beta} - \beta_0)}{\sqrt{\hat{V}(\hat{\beta})}} \tag{4.7}$$

The z-score is the input for the calculation for the p-value. The p-value is the smallest level at which H_0 can be rejected. The p-value is determined using a standard normal table. The standard normal table shows the values of the cumulative distribution function of the normal distribution. Therefore, the table is symmetric around the mean.

To calculate the final p-value the standard normal table needs the z-sore and the significance level as an input parameter. The significance level is denoted by α and is traditionally 1% or 5%. The p-value can directly be compared to the significance level. If the absolute p-value is lower than the significance level, the null hypothesis is rejected and a dependency between the according coefficient and the dependent variable is proven. [48]

$$reject \quad H_0 \quad if \quad and \quad only \quad if \quad \alpha > p - value \tag{4.8}$$

The aim of this thesis is to find out if airline attributes have a significant influence on the market share. The p-value gives an indication about the significance of each coefficient. If the null hypothesis is true, there is a very strong evidence that there is no influence of the coefficient on the dependent variable. The non-significant parameters are deleted from the model one by one. Deleting variables have impact on the results of the model and specifically on the β-coefficients. In the end, when only the significant variables are left over, the final output of the market share model is obtained.

4.5. Correlation

Before the GLM calculation can be performed, a correlation analysis has to be executed for every set of independent variables. In the correlation matrix the correlation coefficients can be found. To detect correlation, a correlation matrix is used to specify if the independent variables show correlations between each other. In the correlation matrix numbers between -1 and 1 can be found. The number indicates to what extent a linear relationship exists. A negative correlation coefficient means that if 'variable one' increases 'variable two' decreases whereas a positive correlation coefficient indicates the opposite. A correlation coefficient of 0 represents no correlation between the two variables and and a correlation coefficient of 1 means that the variables are perfectly collinear which means that the variables are linked with an exact linear relationship. [9]

The statistic software STATA® 13 is used for this project which is also able to calculate the correlation matrix. It shows an estimation for the product-moment correlation coefficient ρ which was popularized by Pearson [33]. In this case the weights w_i are equal to 1 so it can be taken out of the equation. $\bar{x}$ and $\bar{y}$ represent the mean value of the independent variables x and y. [41]

$$\hat{\rho} = \frac{\sum_{i=1}^{n} w_i(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} w_i(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n} w_i(y_i - \bar{y})^2}} \tag{4.9}$$

If the correlation coefficient is higher than 0.7 the two variables are assumed to be highly correlated according to Hinkle et al. [19]. If two variables are highly correlated, such a strong interdependency is present that the variable does not only represent the airline attribute that is stands for. Therefore, one of the two variables has to be deleted from the model. The remaining variable also represents the deleted variable. Finally, the GLM calculation can be performed with the remaining set of independent variables to generate meaningful results.

4.6. Goodness of fit

The goodness of fit is used to determine how good the proposed model actually fits the data that was the input for the model. Three different measurements can be used which are the pseudo R^2, deviance and the Pearson's goodness of fit. The theory behind those models are presented in Section 4.6.1, 4.6.2 and 4.6.3 respectively. All theories are only valid for relative comparison. This means that it cannot be determined of the model withs the data well, it can only be determined with model with the same input data fits the data better.

4.6.1. Pseudo R^2

To measure the goodness of fit, the pseudo R^2 can be used. To be more specific, the McFadden's pseudo R^2 is used in the GLM calculation. The definition of this pseudo R^2 can be found in 4.10. In the equation, $\hat{l}$ stands for the estimated likelihood, M_{Full} for the model with predictors and $M_{Intercept}$ for the model without predictors. Another term for the model $M_{Intercept}$ is null model which is a model that tries to predict the dependent variable without using the independent variable UCLA: Statistical Consulting Group. [46]

$$R^2 = 1 - \frac{ln\hat{l}(M_{Full})}{ln\hat{l}(M_{Intercept})} \tag{4.10}$$

The equation indicates how much improvement can be gained from the full model in comparison to the intercept model. Therefore, a high value of the pseudo R^2 represents a better fit of the model than a low value. However, it has to be pointed out that this is only true for models with the same input data. So different sets of independent variables can be tested, but not different markets or liberalization situation.

4.6.2. Deviance

The deviance one measurement of the goodness of fit of the model. It is defined as twice the difference between the likelihood of the saturated model and the likelihood of the model estimate and is shown in Equation 4.11. The saturated model includes the maximum possible number of parameters with no redundancies which means that there are as many observations as number of parameters. Following from this fact, the vector of parameter ψ is similar to the earlier defined parameter η with the difference that ψ is defined for the maximum number of parameters and η is defined for the number of parameters that need to be estimated. $l_s(\hat{\psi})$ is then the likelihood of the saturated model and $l_i(\hat{\beta})$ is the likelihood of the model estimate which is defined earlier in Equation 4.3.

$$D = 2\left[l_s(\hat{\psi}) - l_i(\hat{\beta})\right] \tag{4.11}$$

With this equation the fit of the link function and the linear predictor is tested for the data available. This means that the deviance tests whether or not the link function and the linear predictor of this model fit the data better than another model. The lower the deviance value the better the fit. Unfortunately, it is not possible to give a specific scale for the deviance only a relative comparison is possible [23].

4.6.3. Pearson's goodness of fit

The Pearson's goodness of fit statistics is named after Karl Pearson who introduced this fitting theory in the publication Pearson [34] in 1900. To make it applicable for GLM, the original Pearson chi-square statistics is converted to the formulation found in Equation 4.12. The numerator includes the squared difference between the dependent variable y_i and the estimated mean function $\hat{\mu}_i$ whereas the denominator shows the variance function which was earlier defined in Equation 4.5 for GLM with robust standard errors and clustering.

$$\chi^2 = \sum \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{4.12}$$

The Pearson's goodness of fit tests the same properties as the deviance but follows a different approach. Whereas the deviance uses the likelihood ratio test, the Pearson's goodness of fit uses what they call the "score test" in Equation 4.12. This results in the fact that an approximation can be established as can be seen in Equation 4.13. Similar to the de [23].

$$\chi^2 \approx D \tag{4.13}$$

4.7. Impact of the independent variable on market share

In this section, it is explained how to calculate the impact of one independent variable of market share. The output of the GLM calculation is a β-coefficient value for each significant attribute. The magnitude of the β-coefficient is dependent on the magnitude of the independent variable. Therefore, it is not possible to compare the β-coefficient directly. This coefficient needs to be reformulated into a parameter that is comparable since it is of interest if the impact of an independent variable has changed due to the liberalization.

To do that a new value that is called $impact_j$ is created where j stands for the independent variable under consideration. The $impact_j$ can be calculated with the help of $range_j$ and the β-coefficient of the corresponding independent variable. The $range_j$ represents the difference between the minimum and the maximum value of the independent variable j. Then the $range_j$ is multiplied by the corresponding β_j and divided by the sum of all ranges times the $beta$-coefficient. The coefficient β_0 stands for the constant that is present in most models.

$$Impact_j = \frac{range_j * \beta_j}{\sum(range_j * \beta_j) + \beta_0} \tag{4.14}$$

The market share can be calculated using Equation 4.2 which represents a s-shaped distribution of the market share. The contribution of β_0 is constant which means that a market share can be calculated that is purely influenced by β_0. In this report, this contribution to market share is also referred to as "base" market share. After that the independent variables contribute. Since the market share is s-shaped distributed a direct contribution in percentage points is not possible. Therefore $impact_j$ is introduced which is a percentage. This percentage of impact can be compared between the different independent variables but also compared for different situations, before and after liberalization.

4.8. Concluding remarks of the market share model

The methodology that is chosen for the market share model is the GLM of the binomial family with logit link function. This method was found to be suitable for market share models since the dependent variable is restricted between 0 and 1 but can take over any value between the boundaries. The software that is chosen is the statistic software STATA® because the GLM methodology is implemented in this software.

Many passenger choice models are created with the logit model. Therefore, it was assumed that this model had potential to be suitable for the market share model too. However, the dependent variables of the logit model can only take over either 0 or1. The GLM is only suitable for fractional dependent variables because the β-coefficients are determined with the quasi-maximum likelihood estimation. Additionally, robust standard errors are introduced because the independent variables do not need to be identically distributed with this addition.

Before the GLM can be performed, the variables need to be tested on correlation. If two variables are highly correlated, one of the two needs to be deleted from the set of input variables. The results of the GLM are β-coefficients for each independent variable. With the p-value it can be determined if an independent variable is significant. The non-significant variables are deleted from the model one by one. To test if the model fits the data, a goodness of fit analysis is performed. These steps need to be taken to create the market share models.

To answer the research questions it is not sufficient to point out the significant variables but it is also needed to specify the impact on market share. This impact value can then be compared between markets, sets of independent variables and situations, before or after liberalization.

5
Case Study

The market share model, presented in Chapter 6.1, is applied to a low-demand market. The Azorean archipelago was chosen to be a suitable low-demand market for the case study because it was liberalized recently.

The chapter consist of several sections. In the beginning a general description is given about the situation on the Azores in Section 5.1. In Section 5.2 the dependent and independent variables are introduced that are used in the market share model. Finally, in Section 5.4, the final model formulation is presented that is used to generate the results presented in the following chapter.

5.1. General description

The Azorean archipelago consists of 9 islands each having a separate airport with a population of about 245000 inhabitants. It is situated approximately 1500 km off the coast of the Portuguese mainland as can be seen in Figure 5.1. The island group belongs to Portugal and therefore to the European Union. Its remote location in the Atlantic Ocean makes it strategically interesting for not only the military but also for commercial air transport. The US maintains a military base on Terceira (TER) island which is open to civil aviation and functions as one of the two main gateways alongside Ponta Delgada (PDL) airport on the main island, São Miguel. Furthermore, the archipelago is important and classified for extended-range twin-engine operational performance standards (ETOPS). Also, the control center for the Santa Maria Oceanic Control Area is located there which is in control of the majority of the traffic between Europe and North America.

SATA is the home carrier of the Azorean archipelago which is fully owned by the Azorean government. Because of its remote location the Azores Islands are dependent on air transport to provide mobility to the inhabitants. Furthermore, the company provides more than 1000 jobs on the islands divided over the different companies SATA group is split into. SATA Air Açores performs all regional flights between the 9 islands whereas SATA International serves the Portuguese mainland (mainly Lisbon (LIS) and Porto (OPO)) as well as other major cities in Europe and North America. There are many Azorean immigrants living in the US and Canada which results in a high Azorean population in cities like Boston, Toronto and Oakland. These cities are also served by SATA.

Figure 5.1: The Azorean archipelago [5]

SATA maintains a fleet of 13 aircraft. The regional network is served by two DHC-8-200 (DH2) with 37 seats and four DHC-8-400 (DH8) with 80 seats which are also used for flights to the Canary Islands. The three Airbus A320-200 are used for flights to the Portuguese mainland and the major European cities whereas the three Airbus A310-300 and the new A330-223 cover flights to North America.

5.1.1. Public Service Obligations before liberalization

The air transport market of the Azorean archipelago was controlled by the PSO's. The European Union (EU) enforced PSO's to various remote cities and areas. The purpose of these laws is to maintain development of the economy on the island or remote area because the EU is also responsible for the infrastructure and supply of the people that are living in those places. The Azores is one of those places in Europe.

There are two types of public service obligations: one where an airline is forced to fly the required routes and if they are making losses on the route, the EU will compensate for it, and the one where the airline is willing to fly the routes the EU asks for. The regional public service obligations for SATA Air Acores for the inter-island flights are enforced but the domestic public service obligations for the flights between the Azores and the Portuguese mainland are not enforced. However, SATA does not fulfill all domestic public service obligations alone. SATA flies in code share with the Portuguese airline TAP to fulfill the requirements together.

Table 5.1: Public service obligations

	Frequency	Seats Summer	Seats Winter	Cargo Summer [ton]	Cargo Winter [ton]
PDL-LIS	1/day	240,000	111,900	14,000	75,000
TER-LIS	4/week	140,000	64,600	8,000	4,400
HOR-LIS	3/week	60,000	28,000	1,000	500
PDL-OPO	2/week	55,000	22,500	-	-
SMA-LIS	1/week	8,100	5,500	-	-
PIX-LIS	1/week	9,500	5,500	30	20
TER-OPO	1/week	-	-	-	-
PDL-FNC	1/week	17,000	5,600	-	-

Not only the seats and the frequency are enforced by the government but also the prices of the route in the domestic and regional public service obligations. These prices are set and are independent on the number of stops between origin and destination.

For this case study only the domestic PSO's are relevant because these regulate the flights between the Portuguese mainland and the Azores. The domestic public service obligations are always valid for one year starting on the 18^{th} of December. The domestic PSO's of 2014 enforced regulation of 7 routes between the Azores and Portugal mainland or Madeira. The characteristics of these regulations can be found in Table 5.1. [15]

5.1.2. After liberalization

Crucial for the development of SATA are the domestic PSO's. On 18^{th} of December 2014 the new domestic PSO's came into force that did not include restrictions on the routes between PDL-LIS, PDL-OPO and TER-LIS anymore, unlike the PSO's that were valid until that point [15]. This means that on these routes the air traffic is liberalized. By liberalizing the routes described, the EU does not guarantee a monopoly position anymore which results in the fact that new competitors entered the market. On the 29^{th} of March 2015, the start of the summer schedule 2016, competitors actually started to fly to the Azores. The competitors on the routes between PDL-LIS and PDL-OPO are low-cost carriers.

On the route LIS-PDL-LIS there are 4 airlines flying after the liberalization, namely SATA, TAP, Ryanair and easyjet. Before liberalization there were only SATA and TAP flying. On the route OPO-PDL-OPO SATA had a monopoly position before liberalization and after liberalization the low-cost airline Ryanair joined the route.

5.2. Variables

In the literature review in Section 2 it was found that the generalized linear model was the most promising methodology to be used for the market share model. The model consists of a dependent variable, that is further explained in Section 5.2.1, independent variables, addressed in Section 5.2.2, and β-coefficients that belong to each of the independent variables.

5.2.1. Dependent variable

The dependent variable of the presented market share model is the market share itself. Market share is defined as the percentage of the total passengers transported in the particular market. In the market share model, the dependent variable is represented by the letter Y as mentioned in Equation 2.6.

In this section, it is explained how historical data referring to the total passengers that are transported by each airline, is obtained and managed. In the first sub-section, it is evaluated how the calibration data is generated and then possible data correction is explained.

Generate calibration data

Data is collected for both situations, before and after the liberalization. The calibration data consists of one observation per day. This means that the market share per day needs to be determined which is the data for the dependent variable. The observation is complete with the value for each independent variable for the corresponding day.

The basis of the calibration data is the data received from the Azorean airline SATA. This passenger data includes all flight related information like the passenger numbers, actual flight times and delays for all flights of SATA and TAP between Ponta Delgada and the Portuguese mainland. The passenger numbers can be used to calculate the market shares for SATA and TAP.

Before liberalization, SATA and TAP were the only airlines flying in the market so the market share can be directly calculated from the passenger flight data. After liberalization, the LCC's entered the market. From the LCC's, no detailed passenger data is available. The only data that is available from Ryanair and easyjet is the total number of passengers that are transported to and from Ponta Delgada for each months. Ryanair and easyjet are joined in the term LCC's from now on. The total passenger data for each market and each airline is made available from the airport reports that are published by the airport in Ponta Delgada every month.

The goal is to find the market share of SATA per day. The passenger numbers of SATA are known so what is missing is the total number of passengers that are transported by all carriers in the market from Ponta Delgada to Lisbon and to Porto. For Lisbon this includes 4 airlines, SATA, TAP, easyjet and Ryanair and for Porto, only SATA and Ryanair are flying. So, the the daily passenger numbers of the LCC's are needed to calculate the total daily passenger number.

Focusing on the route between Ponta Delgada and Lisbon first, the daily passenger number of the LCC's can be calculated using the total number of passengers that are handled at the airport of Ponta Delgada per day and subtracting the number of passengers that are not transported in either of the markets under consideration. This total number of passengers is given in the airport report in a graph. To calculate the exact number, the number of pixels are counted. The total number of transported passengers in that month is known, it can be calculated how many passengers are represented by one pixel. If this is known, the total number of passengers handled by Ponta Delgada airport, is calculated. A schematic visualization can be seen in Figure 5.2, in which the different airlines (and routes) can be seen that contribute to the total number of passengers handled at the airport. That the figure visualizes daily passengers numbers.

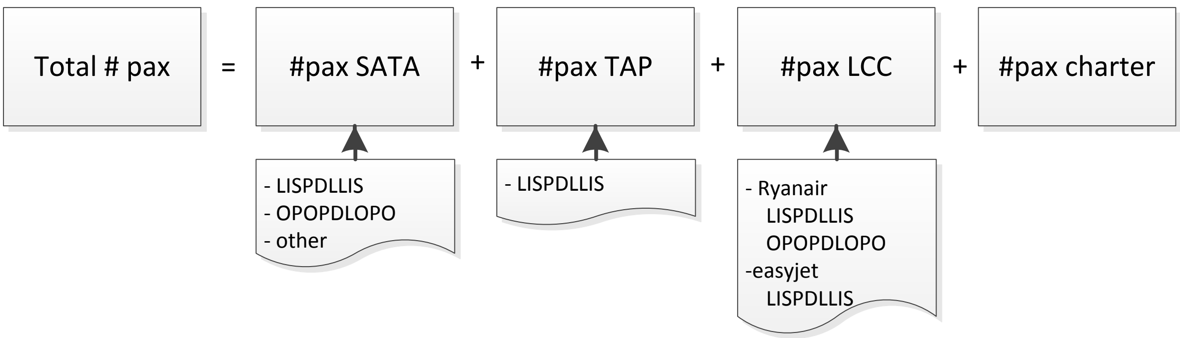

Figure 5.2: Visualization of the calibration data

There are 3 parts that need to be subtracted from the total number of passengers that are handled by Ponta Delgada airport to reach to the number of passengers that are transported by the LCC from Ponta Delgada to Lisbon. These include: 1. the number of passengers that are transported by SATA in different markets than the one under consideration; 2. the number of passengers that are transported by other airlines to and from different destinations than the under consideration; 3. the number of passengers that are transported by Ryanair to Porto.

The passenger number for point 1 is known from the passenger data provided by SATA. The numbers for point 2 are mainly passengers that are transported by European charter airlines like Jetair, Arkefly (now renamed to Tui Fly) and Primera Air. Air Berlin offers regular weekly flights from Düsseldorf to Ponta Delgada. It is not officially a charter airline but for sake of convenience these passengers are counted as a charter airline. The schedule of the charter flights is provided by SATA that has access to the schedule via the ground handling provides at the airport in Ponta Delgada. The monthly passenger numbers are taken from the airport report that lists the passenger transported by origin/destination. The passenger numbers are spread evenly over the number of flights that are provided by the airline each month and then assigned to the corresponding day of the flight.

The final thing that is missing is daily passenger numbers of the LCC's in the market not under consideration. The focus is on the market to Lisbon which means that the passenger flying to Porto with Ryanair need to be substracted. The monthly passenger numbers are available and it is known that Ryanair offers one return flight per day except on Saturdays. Whereas the flight of the charter are spread evenly over flight flights because they are usually weekly flights, this is not accurate enough for the flight from Ryanair between Porto and Ponta Delgada. Weekly seasonality is present. This seasonality is accounted for by finding a factor for every weekday with which the average passenger number can be multiplied. It is assumed that the passengers that use SATA will be representative for the passengers that use Ryanair on the route between Ponta Delgada and Porto. The SATA data is available, so for every month, a factor is calculated for every weekday that can be used to approximate the weekly seasonality.

Now finally, all components are known that are needed to calculate SATA's daily market share. The daily passenger numbers of SATA can be divided by the daily total number of passengers that are transported in the market under consideration. The market share of SATA is used as dependent variable in the GLM calculation. Every observation consists of values for the dependent and independent variables. How the data for the corresponding independent variables are selected is described in Section 5.2.2.

Data correction

Unfortunately, during the calibration of the data some problems occur. When checking if the calculated number of passengers transported by Ryanair between Lisbon and Ponta Delgada actually fits in the number of aircraft for the given frequency it shows that on some dates there are too many passengers for the flights. Also it is suspicious that when this phenomenon occurs, the previous or following day shows passenger numbers that are rather low. The first thought was, that there were flights missing in the calculation because this phenomenon occurs periodically. However, after looking into the flight details it turns out that some flights are delayed until after 24:00 o'clock so the passengers are counted for the following day in the airport report but for the previous day in the counting of the airline. The weather on the Azorean archipelago is changing quickly and the weather conditions are often too rough to land an aircraft which means that they have to divert or be delayed. This is the reason why the passenger numbers do not fit at some days.

To solve the problem it was necessary to go through all the flights that were made in the period between April and July 2015 and remove the passenger numbers from the day they were handled at to the day they were scheduled. Ponta Delgada airport can facilitate 620 arrival and 556 departing passengers per hour according to Sandro Raposo, the director of the network and revenue department at SATA. If a flight had a block-on time from just before 24:00 o'clock these passenger flow numbers were used to calculate the passengers that are counted on the wrong date and needed to be moved to the scheduled date.

The feasibility data correction was later verified by comparing the calculated and the actual number of the low-cost carriers per month. On the route between Porto and Ponta Delgada this number is the same since the passenger numbers are divided over the month with taking the weekly seasonality into account. On the route LIS-PDL-LIS the calibration data is calculated. The passenger numbers for SATA and for TAP are taken from the real flight details and correspond to the number that are used in the calculation. The passenger data of the LCC's relies on approximation which results in differences between the real and the calculated data. The differences between the passenger data are spread between 1.27 percentage points in April and up to 4.68 percentage points in June. Passenger number differences until 5 percentage points are assumed to provide a reasonable fit which is met for the described calculation.

5.2.2. Independent variables

The independent variables represent the airline attributes in the logit model calculation. It is denoted with the letter X_i in equation 2.4. The independent variables describe the airline characteristics under which the passengers book their tickets. Typical airline attributes are ticket price and frequency but also other attributes are considered in the attribute selection which is explained in this section. It is also shown how the data for the attributes is collected and if the attributes are correlated.

Attribute selection

First of all, the attributes that are relevant on the routes between the Azorean archipelago and the Portuguese mainland need to be selected. A literature study was conducted and an overview was created about which airline attributes are used in which paper. This overview, sorted by the name of the first author, can be found in Table 5.2.

At first glance it can be seen that the ticket price is the attribute that is selected from almost all the authors. Ticket price is one of the attributes that are intuitively selected when talking about passenger booking behavior because it is commonly a driving attribute for passengers. To distinguish the different booking behaviors, the ticket price on two different time points is selected. One time point is the ticket price two months before departure and the second time point is one day prior departure. The low-cost carriers tend to be cheap two months before the departure of the flights but raise their fares closer to the departure date. They do that so that people do not have the variety of choices and are forced to book for a more expensive price. The ticket prices of SATA are relatively stable the months before the departure. However, a week before departure the company often lowers the ticket price to fill up flights that have seats left over.

The attribute that is chosen almost as often as the ticket price is the frequency that an airline provides flights on a specific route. If an airline flies more frequently, the passengers are more flexible with their flight choices. About half of the papers choose to take the natural logarithm of the frequency as an attribute. The first to introduce the natural logarithm on the frequency in airline passenger choice behavior is Ben-Akiva and Lerman [7]. It is argued that the natural logarithm can be taken because of diminishing returns that are expected in airline attractiveness from increasing the value for the frequency high enough. Furthermore, the different flights can be seen as different alternatives for the passengers to choose from. So if the frequency increases above 1, the choice is of the aggregated kind. In this

Table 5.2: Attribute selection

Year	Authors	Frequency	Ticket price	Service quality	Total trip time	trip length	# airports	safety records	aircraft size	available seats	time of day	comfort	food	on-time performance	FF member
2001	Adler	1	1												
2010	Adler	ln	1		1										
2014	Adler, Fu, Oum, Yu	ln	1		1										
1992	Alamdari, Black	1	1	1											
2008	Carrier		1							1	1				
2008	Concepion, Espino	1	1									1	1		
2003	Hsu, Wen		1		1										1
2005	Hsu, Wen	1	1												
2010	Levy, Panou	ln	1		1										
2006	Park, Robertson, Wu			1											
2001	Pels et al.	ln	1												
1995	Proussaloglou	1	1	1		1							1	1	
2001	Suzuki	1	1	1		1	1	1							
2005	Wei, Hansen	ln	1						ln	ln					
2007	Wei, Hansen	ln	1						ln						
2010	Wen, Lai	1	1	1							1			1	
2011	Wu et al.				1										
	Total	13	15	5	5	2	1	1	2	2	2	2	2	2	1

thesis, both options are analyzed. Since the frequency in this market varies only between 0-6 per day there are little diminishing returns expected. This is the reason why the frequency attribute is chosen without the natural logarithm for the market share model.

Additionally, the frequency per week was introduced to the model. When leisure passengers visit the island one or two weeks are a typical period the passengers choose for their vacation. This means that if an airline wants to capture these passengers, a periodical weekly frequency has to be offered. Also, a passenger will rather choose for an airline if a variety of return flights is offered. This makes planning more flexible. That is the reason why this attribute is chosen to be considered in the market share model too.

Another attribute which is expected to have significant impact on the passenger choice behavior is the time of the day the flight is departing. For business passengers it is important that they do not waste time before and after meetings and leisure passengers want to get the maximum out of their vacations. Often passengers are also willing to pay an extra amount of money if the flight leaves at an convenient hour. The departure time of the flights in the market are readily available which means that it is possible to use the attribute.

There are time slots identified that have a higher value than other times. These time slots are between 7:00 - 9:00 o'clock in the morning and 17:00 and 19:00 o'clock in the afternoon which are based on interview with Sandro Raposo, the network and pricing director of SATA. If an airline has a flight which has the departure time in this time slot, the airline is awarded with one point. The points are added together for the attribute that is called time points.

The next attribute that is chosen is on-time performance. This is defined as the ratio of flights that are on-time over the total number of flights. This performance value is usually published by the airline on a monthly basis. An aircraft is defined to be late if the delay is at least equal to 15 minutes Prince and Simon [36]. Therefore, a flight is assumed to be on-time if its delay is maximum 15 minutes. On-time performance is a service variable that represents the reliability of the airline.

To capture the market changes over the months, monthly dummy variables are introduced for every month for which observations are available. This dummy variable indicated the change in market share compared to the base months which is chosen to be April. This variable should capture seasonality and the short-term changes that are expected due to liberalization. Especially because airlines could have sold seats even before liberalization was announced.

These are the independent variables that are selected. When the GLM calculation is performed, a constant is added to the market share model. This constant represents all effects that are not included in the independent variables. Examples of these effects could be other service variables like free food on-board of the plane or if they are charged of taking the first and second piece of hold-luggage. Furthermore, comfort effect like the seat pitch or the availability of a business class could be the reason why passenger choose for a certain airline. Additionally, if there were major safety concerns with one airline recently, a passenger is likely to bypass this airline.

Another effect that is captured by the constant is the loyalty effect. Since SATA is an Azorean airline, the inhabitants of the Azores are expected to have a preference to fly with SATA. SATA is also a big employer on the archipelagos which could be a reason for loyalty. Also, legacy airlines like to create loyalty with frequent-flyer programs. Furthermore, if a passenger wants to fly from the Azores to Lisbon or Porto but this is not their final destination the number of airports in the network is one of the attributes that have significant impact on the choice behavior of the passengers. The passengers are forced to book a legacy carrier because LCC's only offer point-to-point transport which would mean that they have to pick-up their luggage and check it in again. Also, the number of airports in the network represents that if a passenger wants to fly to a certain destination and only one airline offers the trip in one flight it is likely that the passenger will choose that airline. TAP is part of the star alliance which will increase the number of airports in the network because then other airlines in the alliance can take over connecting flights. SATA is not part of such a global alliance but works closely together with Virgin America which increases their destinations in North America.

Data collection

The data collection for the attributes that are chosen in the previous subsection is crucial for the development of the project. If no data is available then the attribute cannot be processed in the market share model.

The data collection of ticket price was the most elaborated. As mentioned before, the ticket price was split in two attributes: ticket price two monthss prior departure and ticket price one day prior departure. SATA provided a data sheet of all the tickets that were sold on a particular day on one of the routes and dates under consideration. Since also any kind of discount ticket and tickets from employees are captured in this overview the data is not fully precise but gives a good indication about the actual ticket price of the flights on that day. The ticket price was calculated by dividing the total income on that day for that route over the number of tickets that were sold on that day for that date and route.

The ticket price of Ryanair and easyjet was collected by checking the prices in the online booking systems. When a flight of easyjet and Ryanair was performed on that day, the ticket price is averaged. Unfortunately, no data was available for the LCC's ticket price two monthss prior departure for the month May. This is solved by taking the average ticket price of, for example, the first Monday of April and the first Monday of June and is then rounded to the nearest typical Ryanair or easyjet price. This way an extensive price database was created to serve as input data for the market share model.

A major help in the attribute data collection were the flight details of all the flights from SATA and from TAP which was made available by SATA. With this information the frequency per day, the time of the day the flight departs and the on-time performance could either be taken directly from the dataset or could be calculated with the help of this data for the airlines TAP and SATA. For Ryanair and easyjet the online booking system was used to determine the frequency per week and the flight time which is used to calculate the time points. The on-time performance data of the LCC's are more problematic because they are published for the complete network per month and not per route. This value is used for the market share model but it should be handled with caution.

SATA offers several flights per day between Ponta Delgada and Lisbon and approximately one direct daily flight from the Ponta Delgada to Porto. TAP offers no flights to Porto but one daily flight on the route LIS-PDL-LIS. Ryanair bought one aircraft which is exclusively used for operations in and out of Ponta Delgada Flightglobal [16]. The daily route follows the following path: PDL-LIS-PDL-OPO-PDL-LIS-PDL. Only on Saturday a return flight to London Standsted is offered after the morning Ponta Delgada to Lisbon turn-around. easyjet is the airline that offers flights from Ponta Delgada to Lisbon on Tuesday, Thursday and Sunday in April and May. In June one weekly flight is added on Saturday. The airline does not operate on the route to Porto.

The monthly dummy variables are independent variables for the months May to July. If a flight is served in the corresponding month, the independent variable will take over the value 1 and is 0 otherwise. The month April is taken as a reference variable which is why it is not an independent variable. This means that the result is always in comparison to the reference month.

In the end, some attributes were not selected to be independent variables because of several reasons. They were expressed in a dummy variable or in a variable that has little variation. Both characteristics are prone to be correlated with other variables. Therefore, it was not possible to include those variables in the market share model directly but their effect can be found in the constant. The constant is a fixed variable that is automatically calculated by STATA®. It can be seen as a correction factor for the regression equation that was introduced in Equation 2.6. However, all attributes that were considered but not selected are also captured in the constant.

From attribute to independent variable

Now, all attributes are selected and the corresponding data is collected. The next step is to think about how to include the data in the market share model. For every airline in the market, one observation is present for every day between April-July 2015 (122 days). This would theoretically result in 366 observations. The observations include the market share as the dependent variable and the corresponding values for each independent variable for each day.

However, if the GLM calculations are made with the dataset as described there is no link between the airlines. This means that the result is one equation that holds for all airlines in the market. This results in the fact that if data of several airlines is included in the equation for the same day, the market share will not add up to exactly 100%. This inaccuracy leads to the following decision: the independent variables are taken as a ratio of SATA divided by its competitors. An advantage of this approach is, that it is possible to capture changes that are in relationship with the competitors. Passengers also compare the airlines in the market so it makes sense to take the ratio of the attributes.

The different attributes are treated differently when creating the independent variable that is used in the market share model. All variables that represent a frequency share are estimated using Equation 5.1, which result in a percentage of SATA against all flights in the markt. The variables that represent frequency shares are $freq_{day}$ (frequency per day), $freq_{week}$ (frequency per week), and $time\_points$ (frequency in a preferred time slot). The remaining variables are estimated using Equation 5.2 with the average value of the competitors in the denominator. The remaining variables are $price_{day}$ (price the day prior departure), $price_{2m}$ (price two monthss prior departure), and on_time (on-time performance). The value of this variable is not bound between 0 and 1 but shows for example how much more the ticket price is in comparison with the competitors.

$$freq = \frac{freq_{SATA}}{\sum freq_{all\_airlines}} \tag{5.1}$$

$$variable = \frac{variable_{SATA}}{\overline{\overline{variable_{competitors}}}} \tag{5.2}$$

Using this approach, only the market share of SATA can be estimated with the resulting equation. The data set changes too since only one observation per day is left over. This makes the dataset collapse to a theoretical value of maximum 122 observations. But this value is not reached because of several reasons: First of all, it was not always possible to determine the ticket price the day prior departure because the flight might be sold out. This results in the fact that sometimes $price_{day}$ cannot be determined. If one value of an independent variable cannot be determined, the entire observation is deleted from the model. This method is called list wise deletion and the effect is explained in more detail in Section 6.3.

Second of all, in the market between Ponta Delgada and Porto after liberalization and Ponta Delgada to Lisbon before liberalization, there are several days per week where only one airline is flying. This monopoly position results in market shares of 0 or 1. The observations of these days are removed from the data set, because changes in the independent variable will not trigger changes in the dependent variable. Therefore, these observations would be misleading. The actual number of observations for each market is given in Section 5.3.

Furthermore, it needs to be discussed why there are only 6 standard independent variables (+3 monthly dummies) left over from the variety of the possible variables shown in Table 5.2. This is only possible when knowing the data set of the variables. Trip length and trip time are not included because in this case study, all airlines depart and land from the same airport. No difference between legacy and low-cost carriers is present.

The number of airports, seat pitch, loyalty and service variables like food, frequent flyer membership have one fixed variable per airline. This results in the fact that correlation is present. The principle of correlation in explained in Section 4.5. It means that one variable can be transformed into another variable so that it does not represent one airline attribute alone. Correlation is (almost) always present but the level of correlation is important that can be measured with the correlation coefficient. A correlation coefficient of 0.7 and above is assumed to be highly correlated, which means that one of the variables has to be deleted from the model Hinkle et al. [19].

It is not possible to measure the influence of every independent variable at once. Therefore, different market share models are created that take different independent variables as input variables. Firstly, the market share model is created with all airline related variables. These include $price_{day}$, $price_{2m}$, $freq_{day}$, $freq_{week}$, $time\_points$ and on_time. This model is considered the standard model on which the remaining models are based on.

Secondly, the previously described market share model is downsized by the variable $freq_{day}$. This variable shows high correlations (above 0.7) with the dependent variable market share. Therefore, the influence of $freq_{day}$ 'overrules' other influences. Which means that when high correlation is present, other variables tend to not be significant in the model. Therefore, this model is created to detect the independent variables that are significant if the strong effect of $freq_{day}$ is not present.

Thirdly, the timely effect needs to be measured in comparison with the standard model. Therefore, the 3 different monthly dummies, May, $June$ and $July$, are included in the market share model. This means that the standard model is extended by 3 independent variables. The monthly dummies will show the timely effect of the variables in relationship with the base month April.

5.3. Descriptive statistics

This section presents the minimum, maximum and mean values of each variable in the different markets. The results of the descriptive statistics for the situation after liberalization can be found in Table 5.3. It can be seen that the value of $freq_{day}$, $freq_{week}$ and $time\_points$ are bound between 0 and 1. It can be seen from the table that on average, SATA achieves the highest market share in the market from Ponta Delgada to Porto even though the values for $price_{2m}$ are relatively high. But it can be explained with the high average frequency shares per day and per week.

In the market between Ponta Delgada and Lisbon the most competition is present. Four different airlines offer flights which results in the fact that SATA does not achieve a high frequency share (in neither of the variables) and market share even though SATA has the most flights in the preferred time window. An explanation for this can be found in the Section 6.

Furthermore, the number of observations are always lower than the theoretical value 122. It is the lowest in the market between Ponta Delgada and Porto. This also results in the low value of observations in the combined market. The reason why the theoretical value is not reached has been explained in the previous section.

The descriptive analysis can also be performed for the data before liberalization in the market between Ponta Delgada and Lisbon. The results can be seen in Table 5.4. When comparing the situation before and after liberalization it can be seen that the value for the ticket price variables are higher in

Table 5.3: Descriptive analysis for the variables in the situation after liberalization

after liberalization									
	LIS-OPO			LIS-PDL-LIS			OPO-PDL-OPO		
	min	max	mean	min	max	mean	min	max	mean
$price_{day}$	0.48	1.8	0.95	0.6	3.9	1.37	0.46	3.08	1.05
$price_{2m}$	0.85	4.5	1.96	0.74	4.37	1.83	1.14	5.22	2.46
$freq_{day}$	0.09	0.53	0.36	0.11	0.55	0.34	0.33	0.67	0.5
$freq_{week}$	0.34	0.44	0.38	0.3	0.39	0.34	0.45	0.57	0.51
$time\_points$	0	0.57	0.42	0	1	0.61	0.33	0.67	0.52
on_time	0.66	0.69	0.67	0.86	0.91	0.88	0.86	0.94	0.89
$market\_share$	0.18	0.57	0.39	0.13	0.62	0.39	0.28	0.65	0.43
# observations	64	64	64	102	102	102	56	56	56

the situation before liberalization. The variables that represent the frequency shares are also higher before liberalization. This means that SATA offers more flights in comparison to the competition before liberalization and is therefore able to charge higher ticket prices too. However, SATA is unable to position as relatively many flights in the preferred time window as after liberalization. From the data it can be concluded that the market share decreases due to liberalization.

Table 5.4: Descriptive analysis for the variables in the situation before liberalization

	before liberalization		
	LIS-PDL-LIS		
	min	max	mean
$price_{day}$	0.45	3.93	1.59
$price_{2m}$	0.46	6.13	2.01
$freq_{day}$	0.6	0.83	0.73
$freq_{week}$	0.78	0.87	0.83
$time\_points$	0	0.5	0.42
on_time	0.57	0.72	0.65
market_share	0.64	0.92	0.77
# observations	71	71	71

5.4. Concluding market share model for the Azorean case study

The GLM is applied to the case study of the recently liberalized Azorean archipelagos. Two routes are analyzed with are the route from Ponta Delgada to Lisbon and to Porto. Also a combined market is added that merges those routes because passengers might choose them interchangeably. The focus is on the impact of liberalization which means that the market between Ponta Delgada and Lisbon is analyzed before and after liberalization. This is not possible for the market to Porto because SATA had a monopoly position before liberalization. Three market share models are developed that differ in the set of independent variables. First, all airline related independent variables are considered. Then $freq_{day}$ is removed from the model and finally the market share model is expanded by monthly dummy variables.

The final model formulation that is used in the following section to generate the results can be found in Equation 5.3. The final selection of the attributes include the ticket price for the flights one day and two monthss prior departure relative to its competitors that is represented by $price_{day}$ and $price_{2m}$ respectively. Furthermore frequency share is included in the model in two versions: $freq_{day}$ and $freq_{week}$. To capture the influence of the time of departure on the market share, the $time\_points$ variable is introduced in the model. Also, on_time represents how reliable the airline is for the passenger. Additionally, the monthly dummies are included in the model if the short-term effect of liberalization is analyzed. Finally, when the GLM calculation is performed a constant is added to the model that is denoted by β_0 in the equation below.

$$\eta = \beta_0 + \beta_1 * price_{day} + \beta_2 * price_{2m} + \beta_3 * freq_{day} + \beta_4 * freq_{week} + \beta_5 * on\_time + \beta_6 * time\_points \tag{5.3}$$

$$marketshare = \frac{exp(\eta)}{1 + exp(\eta)} \tag{5.4}$$

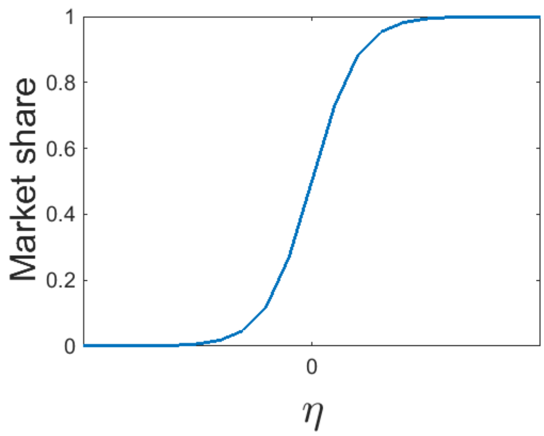

Figure 5.3: S-shape of market share vs. η

Equation 5.3 is the equation that is needed to determine the score η. η is then used in the market share equation that can be found in Equation 5.4. The market share equation represents the typical S-shape curve that can be found in Figure 5.3. It is not only true for frequency shares but also for the GLM. It represents that if all variables A GLM is performed to evaluate the value of the β-coefficients. These coefficients give an indication about the impact of each variable and about the change in impact due to liberalization. Now, all components are known that are needed to generate the results.

6 Results

In the previous chapters the methodology of the market share model is described and the case study is introduced. In this chapter the results of the market share model are presented. First the β-coefficients of each case are presented and then the impact of each significant variable is analyzed.

Before the β-coefficients can be determined, a correlation analysis is performed. For this the correlation matrix is calculated. The independent variables that have a correlation coefficient of more than 0.7 are deleted from the model [19]. The correlation matrix of each case can be found in Appendix A.

The chapter is divided in several sections. Section 6.1 shows the value of the significant variables that where found using the market share model. The section is structured to give the results of all market share model that are created for all markets in the case study. In Section 6.2, the impact of the significant variables on market share are given. In the next section, Section 6.3, it is shown how the deletion of observation has influence on the results. After that, the fit of the model is discussed in Section 6.4. In the final section 6.5 will give a summary about the results of the market share model.

6.1. Market share model

In this section the significant variables are presented with their corresponding β-coefficients, which is the output of the GLM calculation described earlier. If a variable is found to be significant, the β-coefficient gives an indication of the impact of that particular variable on market share.

First, the situation after liberalization is considered. For this situation, the results of the market share model is presented with 3 different sets of independent variables. First, the model is run for all independent variables that are airline-related in Section 6.1.1. Second, the independent variable $freq_{day}$ is removed from the model and the results are given in Section 6.1.2. Third, in Section 6.1.3, the market share model is extended by monthly dummy variables.

After that, the results of the situation before liberalization are presented. Similarly to the previous, the market share model is given all independent variables as input in Section 6.1.4, and then the market share model is generated without the variable $freq_{day}$ are given in Section 6.1.5. The monthly dummy variables are not present in the model before liberalization.

6.1.1. Results of the market share model after liberalization including all airline-related independent variables

In this section the results of the market share model for the situation after liberalization with all airline-related independent variables are presented in Table 6.1. The full output of STATA for the GLM calculation for this situation and for all the other situations can be found in Appendix B. The following independent variables are considered in the model: ticket price the day prior departure and two months prior departure, frequency per day and per week as well as time points.

On-time performance is another airline-related independent variable which is not regarded in this calculation. The reason for this is, that with the variable the results cannot be validated. *On-time performance* changes throughout the year and because the model is calibrated with only four months of data, the data of this variable is not representative for the full year.

Table 6.1: Results of the market share model after liberalization with all airline-related variables

after liberalization					
	LIS-OPO	LIS-PDL-LIS	LIS-PDL	PDL-LIS	OPO-PDL-OPO
$price_{day}$		-0.13			
$price_{2m}$					
$freq_{day}$	4.43	4.67	5.06	4.50	3.19
$freq_{week}$					-2.47
$time\_points$					
_const	-2.09	-1.90	-2.16	-2.01	

The results of this market share model is structured by markets. First, the significant variables of the joined market between Ponta Delgada and Lisbon as well as Porto is shown in Section 6.1.1. Then, the route LIS-PDL-LIS is looked in Section 6.1.1. In Section 6.1.1, only the flight between Lisbon and Ponta Delgada is considered. After that, the significant variables of only the flight in the opposite direction is shown in Section 6.1.1. Additionally, the market between Ponta Delgada and Porto is considered where the results of the combined return flight is shown first in Section 6.1.1.

LIS-OPO

The first market that is considered, is the combined market of the flights between Lisbon as well as Porto to Ponta Delgada. The data is merged in such a way so that there is one observation per day. This case is considered because the routes are often used interchangeable. This means that if for example, if the flight between Lisbon and Ponta Delgada is sold out, too expensive, or departs at an inconvenient time, the passengers may choose to fly via Porto to Ponta Delgada if that is more convenient for them.

The results in Table 6.1 show that only the $freq_{day}$ variable is a significant variable. All other independent variables are found to be insignificant. Additionally, a constant is added to the model that is also significant. The constant represents all the airline attribute that could have influence on the model but are not specifically included in the model. The constant is negative, which means that if all the variables would be set to zero, SATA would have a market share of less than 50%.

$Freq_{day}$ is correlated with market share which is the reason why no other variable is significant in this case, as can be seen in the correlation matrix in Appendix A. If a variable is highly correlated it tends to 'push' other variables 'out' which means that other variables are not significant. Therefore, another case of the model without the variable $freq_{day}$ is presented in Section 6.1.2. The reason for the high correlation lies in the fact that when an airline offers more flights per day, more passengers have the chance to be transported and the possible market share increases. Since market share is determined per day, the frequency per day has direct influence on the market share of an airline.

It is surprising to see that none of the ticket price variables are significant in the model. This means that no significant relationship between market share and ticket price could be found. When analyzing the data a reason for this can be found.

In Figure 6.1 one can see the market share vs. the two ticket price variables. The independent variables that represents $price_{day}$ can be seen in blue, whereas $price_{2m}$ can be found in red. Also the corresponding linear regression lines can be seen in the figure. It can be seen that neither of the variables follow a clear pattern. Especially $price_{2m}$ is scattered randomly so that no link between the independent variable and the dependent variable can be drawn.

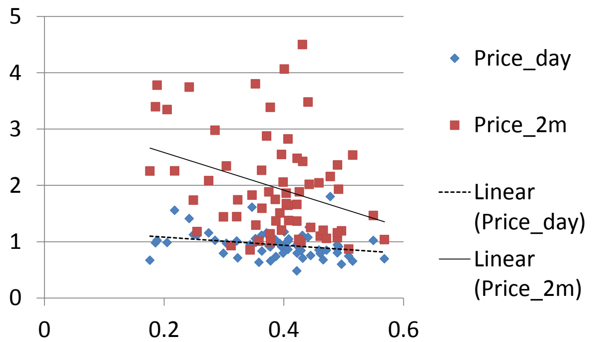

Figure 6.1: Relationship between raw data of ticket price variables and market share

Additionally, neither $freq_{week}$ nor $timepoints$ are significant variables. This means that it seems that it is not important for the passengers to have continuous flights or flights in a specific time window. The reason for is could be that SATA offers at least one daily return flight between Lisbon as well as Porto to Ponta Delgada and back but usually even more flights are offered. Therefore, the passengers are less concerned about their return flight. Higher $freq_{week}$ and $timepoints$ means more flexibility for the passenger but in this case, the variables have no significant influence on market share.

LIS-PDL-LIS

For the route between Ponta Delgada and Lisbon, the flights of four airlines are taken into consideration which are namely: SATA, TAP, Ryanair and easyjet. The data of the LCC's is combined because of the data availability. The data for the dependent and independent variables, that is fed into the statistic software STATA®, is detailedly explained in Section 5. The data is used for the market share model to generate β-coefficients, which are the results that can be found in the corresponding column in Figure 6.1.

Now. it is time to focus on the values of the β-coefficients. In this market the variables $price_{day}$ and $freq_{day}$ are significant. Also a significant constant is added to the model. First of all, it needs to be tested if the sign of the coefficient is represented the meaning of the attribute. The β-coefficient for the attribute that represents the $price_{day}$ is negative. Since ticket price is a cost to the passenger and the attractiveness of the airline will decrease with an increasing ticket price, the sign of the coefficient makes sense. If an airline offers a higher $freq_{day}$ the passengers are expected to change their airline choice behavior in favor for that airline, meaning that the sign of the coefficient also makes sense.

The independent variable $price_{day}$ is significant in the market share model. The mean of this values is equal to 1.37 which means than on average the ticket price the day prior departure is 37% higher than the competitors ticket price. When adding a trend line to the $price_{day}$ data one can see that the ticket price is decreasing when approaching the summer month. All airlines increase their ticket price towards the summer months but TAP increased them more than the other carriers in the market. Therefore, the value of $price_{day}$ drops. Furthermore, Azorean passengers tend to book their flights close to the departure date [38]. It seems that with the liberalization, the passengers base their airline choice also on ticket price the day prior departure.

$Freq_{day}$ is the remaining significant variable in the model. The reason why the variable is significant is similar to the reason in the previous case. Higher frequency on a particular day leads to higher market share at that particular day. The variables are correlated as can be seen in Appendix A. The constant that is present in the model is less negative than in the previous case.

One of the goals of this research is to create an equation that reflects the market and can be used to calculate the market share in the future with the according airline attributes. The equation is of the form as shown in Equation 5.3. It can be seen that a constant value is added to the equation so that the market share can be calculated properly with the significant attributes that are presented.

The resulting β-coefficients of the market share model for the return flight between Ponta Delgada and Lisbon for the situation post-liberalization can be put into a equation that calculates η which can be seen in Equation 6.1. η can then be used to calculate the market share with Equation 4.2.

$$\eta = -0.13 * price_{day} + 4.67 * freq_{day} - 1.9 \tag{6.1}$$

LIS-PDL

In this section the market between Lisbon and Ponta Delgada is considered. The results of this model can be used to to draw conclusions about the difference in passenger behavior when flying to the Azores and back. Therefore, the return flight is split into its two legs and analyzed separately. In Table 6.1 the significant output of the market share model for this market segment can be seen. The correlation matrix of this case can be found in Appendix A along with all the other correlation matrices.

The significant variable in the model is $freq_{day}$ and a constant is added. Because of the high correlation between $freq_{day}$ and the market share of more than 80%, as can be seen in the correlation matrix in Appendix A, no other variable is significant. This is different to the LIS-PDL-LIS model where $price_{day}$ is also significant. The β-coefficient of $freq_{day}$ is positive which is as expected. Since the constant contributes less to the model than in the full return case, the β-coefficient for $freq_{day}$ needs to be higher to achieve a comparable mean market share, which is the case for this model.

PDL-LIS

In this section the flight from Ponta Delgada to Lisbon is considered for the situation after liberalization. It is part of the combined return flight case LIS-PDL-LIS and has the same airlines flying in the market as the previous described market segments.

The results in Table 6.1 it can be seen that only $freq_{day}$ is a significant variable. This attribute can also be found in the LIS-PDL-LIS model and the reason for the stand-alone position is the correlation of $freq_{day}$ with market share. The coefficient for $freq_{day}$ is slightly lower than in the LIS-PDL case. The reason for this lies in the fact that the mean value for $freq_{day}$ is slightly higher in this case compared to the LIS-PDL case.The sign of the β-coefficient makes sense.

Similar to the previous models, a (negative) constant is added to the market share model so that the formula with the airline attributes and the β-coefficients can represent market share correctly.

OPO-PDL-OPO

In this section the results of the market between Ponta Delgada and Porto are given. The β-coefficient for the model can be found in Table 6.1 in the last column.

The first thing that can be noticed is that the the model does not have a significant constant. The significant variables in this case are $freq_{day}$ and $freq_{week}$. That $freq_{day}$ is significant is in line with the results of the previous cases, however $freq_{week}$ is not significant in any other model.

Also, the sign of the β-coefficient is not as expected. If an airline offers more flights per week one would expect a positive contribution to market share, however the coefficient for $freq_{week}$ is negative. The reason for this lies in the fact that there is not constant present. $Freq_{week}$ captures the influences on market share that are included in the constant in the previous models. The constant is negative in all models which is the reason why the variables takes over a negative sign too.

Another reason for the unexpected sign of the $freq_{week}$-coefficient can also be explained with the seasonality of the Azorean market. SATA increases their frequency per week from approximately 10 in the month April and May to 14 in the summer month July to cope with the increasing demand. Ryanair's frequency per week is constant whereas the frequency per week of SATA increases. The market share on the other hand increases in favor of Ryanair over the months. Therefore, it is concluded that an increasing number of frequencies per week is counteracting an increasing market share.

The problem with this model is the data. As explained earlier, the observations are deleted where a value for one of the variables is missing. On the route between Ponta Delgada and Porto only SATA and Ryanair are flying. Ryanair and SATA do not offer daily flights every month which means that on those days, one of the airlines has a monopoly position and the observation is removed from the model. Also, it happened that a flight is sold out or the ticket price value could not be determined. Then the observation is deleted too. In the end, there were only 57 left for the market share model. Less data leads to a lower accuracy of the model. This is another the reason why the variables have unexpected β-coefficients. Because of the problems in the data, the separate routes of the model are not analyzed in detail.

6.1.2. Results of the market share model after liberalization without $freq_{day}$ as independent variable

There is a problem with the independent variable $freq_{day}$ because it is prone to be correlated with market share. To check the correlation, a correlation matrix is created where the correlation relationship between all the variables can be found. The correlation matrices of each case can be found in Appendix A. In most cases the correlation between $freq_{day}$ and market share reaches a correlation above 0.7 which was defined to be highly correlated by Hinkle et al. [19]. Therefore, new results are given in Table 6.2 for each case where the variable $freq_{day}$ is disregarded.

The purpose of this section is to analyze which independent variables have significant influence on market share apart from $freq_{day}$. When an independent variable is strongly correlated with a dependent variable, it tends to "push out" other variables. Therefore, the GLM calculation is performed on the model without this independent variable.

When performing the correlation analysis, it can be found that neither of the remaining variables have a correlation coefficient of above 70%. This can also be seen in the correlation matrices in Appendix A. Therefore, no other variable has to be deleted from the model.

Table 6.2: Results of the market share model after liberalization without $freq_{day}$

	after liberalization				
	LIS-OPO	LIS-PDL-LIS	LIS-PDL	PDL-LIS	OPO-PDL-OPO
$price_{day}$		-0.25			
$price_{2m}$	-0.13				
$freq_{week}$					-3.04
$time\_points$	0.65	0.80		0.57	
on_time					
_const	-0.49	-0.61		-0.58	1.27

The structure of this section is similar to the structure of the previous section. First the combined model is presented where the cases of LIS-PDL-LIS and OPO-PDL-OPO are merged to the model that is called LIS-OPO in Section 6.1.2. Then, the market share model of the market between Lisbon and Ponta Delgada and back is presented in Section 6.1.2. In the following two Sections 6.1.2 and Section 6.1.2, the market is split in the separate flights LIS-PDL and PDL-LIS respectively. Finally, the results for the market between Ponta Delgada and Porto are shown in Section 6.1.2

LIS-OPO

In this section, the combined model of the flights from Ponta Delgada to Lisbon as well as Porto for the situation after liberalization is analyzed. The results of the GLM calculation without the independent variables $freq_{day}$ can be found in Table 6.2.

It can be seen that the independent variables $price_{2m}$ and $time\_points$ are significant in the model. Also a significant constant is added to the model. The value for the constant is lower than the values found the the case when $freq_{day}$ was considered too.

The sign of the β-coefficient of $price_{2m}$ is negative which makes sense because one would expect decreasing market share with increasing ticket price. Furthermore, the sign of $time\_points$ is positive. The variable $time\_points$ represents the frequency share of the flights that depart is the time window between 7-9 o'clock and 17-19 o'clock daily. If the frequency share increases, the market share is expected to increase too. Therefore, the sign of the variable makes sense.

The input data of the model supports the sign of $price_{2m}$. When plotting $price_{2m}$ against market share a decreasing trend line can be found which means that with increasing market share the value for $price_{2m}$ increases. This corresponds with the expectations described above.

A similar analysis is performed for the independent variable $time\_points$. Here, an increasing trend in $time\_points$ can be found for increasing market share. If more flights are offered in the preferred time windows, passengers tend to make their booking with the airline that offers flights in these time windows.

LIS-PDL-LIS

In Table 6.2, the resulting β-coefficients can be found for the market share model in market LIS-PDL-LIS after liberalization without the independent variable $freq_{day}$. It can be seen that the variables $price_{day}$ and $time\_points$ are the significant independent variables in the model. A significant constant is also added to this model. The impact of the constant on market share equals to 35 percentage points.

The independent variable $price_{day}$ has a negative sign and was already significant in the case when $freq_{day}$ was not deleted from the model. When looking at the raw input data for the model one can see a negative relationship between $price_{day}$ and market share. This means that the market share decreases when the value for $price_{day}$ drops. It makes sense that passengers would choose for the cheaper airline when all the other airline attributes are the same.
The independent variable $time\_points$ has a positive sign. This also matches with the expectations because one would expect the market share to increase when more flights are offered in the preferred time window. The mean value for the variable $time\_points$ equals to 0.61 which is a rather high value compared to other models. This also explains the rather high value for the β-coefficient.

Additionally, the correlation coefficient of the variables $freq_{day}$ and $time\_points$ is 58% as can be seen in the correlation matrix in Appendix A. This value is not as high such that one of the variables are deleted from the model, but it is a good explanation why $time\_points$ is a significant variable when $freq_{day}$ is disregarded from the market share model.

LIS-PDL

For the route between Lisbon and Ponta Delgada in the situation after liberalization, no other independent variables is significant as can be seen in Table 6.2. In Section 6.1.1, it is shown that $freq_{day}$ is significant in the model, but when removing this independent variable, it can be seen that none of the other variables have significant influence on the market share.

PDL-LIS

In this section, the results of the market between Ponta Delgada and Lisbon is analyzed for the case when the independent variable $freq_{day}$ is disregarded from the model. The β-coefficient of the only significant variable $time\_points$ can be found in Table 6.2. In this table it can be seen, that there is also a significant constant present in the model. SATA achieves a market share of 36 %. Considering there are actually four airlines flying in that market (of which two are merged to "LCC's"), SATA achieves a relatively high market share when all variables are set to zero. The constant includes all effects that are not measured in the independent variables. Loyalty might be a significant part of that which would explain why SATA might be favorable for Azorean passengers.

The coefficient of the independent significant variable $time\_points$ has a positive sign which corresponds to the expectations as described in the previous section. An increase in market share is expected when the airline SATA offers more flights in the preferred time windows. This positive behavior can also be found when the market share and the $time\_points$ input data is plotted against each other.

OPO-PDL-OPO

The last market that is analyzed in this section is the market between Porto and Ponta Delgada. The market share model is created with all the airline-related independent variables except the variable $freq_{day}$. This way, it is possible to analyze the market on influences apart from $freq_{day}$. Since $freq_{day}$ shows strong correlations with the market share and it Therefore, dominates the influences, it is not always possible to detect these influences in the market.

As can be seen in Table 6.2, the significant independent variable is $freq_{week}$ and a positive significant constant is added to the model. The β-coefficient for the variable $freq_{week}$ is negative which contradicts with expectations. One would expect the market share to increase when the frequency per week is increased too.

A positive constant represents a market share of above 50% when all variables are set to zero. In this case, the contribution of the constant on market share is rather high. The market share will then be decreased by the contribution of $freq_{week}$ according to the market share model.

When analyzing the input data of the model one can see that, the input data of variable $freq_{week}$ also decreases with increasing market share. A reason for this could be that SATA increases its frequencies approaching the summer which increases the value for $freq_{week}$. However, SATA fails to score higher market shares in those months.

Another reason for the unexpected behavior of the model is the number of observations which is already mentioned in Section 6.1.1. With only 57 observations, is it not always possible to create a market share model that generates satisfying results.

6.1.3. Results of the market share model after liberalization with monthly dummy variables

In this section the results of the market share model for the situation after liberalization are given where monthly dummy variables are added to the model. The results of the market share model with monthly dummy variables can be found in Table 6.3. It can be seen that only the return markets are shown in the table. For this extended market share model, the focus lies on the full return markets.

For the monthly dummy variables, the month April is taken as a reference months. May, June and Juli are dummy variables at are added to the market share model with all airline-related independent variables are considered. Therefore, the effect of time and seasonality can be analyzed in comparison with the month April.

Analyzing the impact of time is especially interesting because SATA already sold tickets for their flights in April-July before the LCC's publicly announced that they were entering the market. The market was liberalized earlier but the LCC's started flying to the Azores with the start of the summer schedule which is on the 29^{th} of March 2015. However, the summer schedule is created by each airline before November because this is the time when the slot conferences take place. SATA did not know for sure if the LCC's intend to enter the market before this point. To know the ticket price of the competition, the tickets of the LCC's have to go on sale which happened in December 2014. Only after that time, SATA was able to know the airline attributes the LCC's would compete with on the market from Ponta Delgada to Lisbon and Porto.

But before December 2014, passengers already booked tickets for a flights of SATA between April-July 2015. At that point the passengers did not know that there would be competition and that they would have another airline choice next to TAP. In this section it is analyzed if this phenomenon has influence on the market share model. Then this market share model can be compared to the market share model where the dummy variables are not considered.

On the first glance it can be seen that the independent variable on_time is correlated in each of the markets. on_time represents the on-time performance of SATA divided by the average on-time performance of its competitors. The independent variable is determined per month which means that it is equal to the same value for the whole month. Therefore, it is prone to be correlated. When doing the correlation analysis one can see that on_time is highly correlated in each of the markets with one of the monthly dummies. on_time is deleted from the model when the correlation coefficient is higher than 0.7 which was defined by Hinkle et al. [19]. The correlation matrix for each market can be found in Appendix A.

Table 6.3: Results of the market share model after liberalization with monthly dummies

	after liberalization		
	LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO
$price_{day}$			-0.15
$price_{2m}$			
$freq_{day}$	4.43	4.88	3.08
$freq_{week}$			
$time\_points$			
on_time	corr	corr	corr
may		0.17	
$june$		0.25	-0.26
$july$		0.40	-0.40
_const	-2.09	-2.34	-1.51

LIS-OPO

Similarly to the previous sections, the combined market between Ponta Delgada and Lisbon as well as Porto is analyzed first. The market share model of all airline-related independent variables is extended by monthly dummy variables and created for the situation after liberalization. As can be seen in Table

6.3, neither of the dummy variables are significant in this model. This means that the results are equivalent to the results of the market share model of all airline-related independent variables which can be found in Section 6.1.1. No monthly effects can be detected.

LIS-PDL-LIS

The second market that is analyzed is the full return market between Ponta Delgada and Lisbon. The β-coefficients of the market share model with monthly dummy variables for the situation after liberalization can be found in Table 6.3. It can be seen, that the independent variable $freq_{day}$ is significant in the model and so are all monthly dummy variables. Also, a negative constant is significant in the model.

The significant monthly dummy variables have increasing β-coefficients when coming closer to the summer months as can be seen in the table. This means that compared to the base month April, the market share of SATA is increasing towards the summer. The difference in β-coefficients between May and June is 0.08 whereas the difference between June and July is even 0.15. The actual impact of each months will be discussed in Section 4.14. An increase in market share towards the summer months makes sense, because the Azores is a market with a heavy increase in demand in the summer months which means that seasonality is present.

The remaining significant independent variable is $freq_{day}$. This variable is also significant in all other models. The number of transported passengers is directly linked to the frequency. Therefore, the significance of $freq_{day}$ as well as the relatively high β-coefficient is expected.

One thing that is surprising is that the independent variable $price_{day}$ is not significant. This variable is significant in both models that are presented in previously. For the market share model where all airline-related independent variables are taken into account, $price_{day}$ achieves to be significant even though $freq_{day}$ is highly correlated with market share. In this model, it seems that the dummy variables "push out" the variable $price_{day}$. The dummy variables with $freq_{day}$ represent the market share better than the independent variables $freq_{day}$ and $price_{day}$.

OPO-PDL-OPO

The market share model that is created with the monthly dummy variables is the full return market between Porto and Ponta Delgada. The resulting β-coefficients of the significant variables $price_{day}$, $freq_{day}$, June and July can be found in Table 6.3.

The monthly dummy variables in this model are not increasing towards the summer months but are decreasing. The month May is not significant, but June and July are. The market share of SATA is lower in June and even lower in July compared to the base month April. A reason for this could be that SATA already sold seats on the flights in April before the entrance of the LCC's in the market was announced. Also, travel agencies book seats on flights well in advance. The agencies sell vacation packages with flight and hotel included in the price. Because the agencies guarantee to buy a significant amount of seats on every flight, low prices can be achieved for the costumer.

Next to that, $freq_{day}$ and $price_{day}$ are significant. The reason for the significance of $freq_{day}$ is mentioned already several times but the reason for the significance of $price_{day}$ lies in the fact that the passengers seem to book their tickets close to departure. The sign of the β-coefficient is negative since the market share decreases when the ticket price the day prior departure increases.

6.1.4. Results of the market share model before liberalization including all airline-related independent variables

In this section the results of the market share model with all airline-related variables before liberalization are presented in Table 6.4. The time before liberalization is defined as time before the 29^{th} of March 2015. For the situation before liberalization, only the market share between Ponta Delgada and Lisbon is analyzed, because for the case between Ponta Delgada and Porto, SATA had a monopoly position.

To be able to compare the situation before and after liberalization, the same time period of the year is selected for the data. This means that for the situation before liberalization data from April to July 2014 is taken for the calibration of the market share model. Before liberalization, only TAP and SATA flew on the route between Ponta Delgada and Lisbon and they fulfilled the PSO requirements together. SATA offered between 2-4 return flights daily whereas TAP only flew 4 return flights per week. The same airline attributes are taken into consideration for the situation before and after liberalization. This makes it possible to compare the cases easily.

Table 6.4: Results of the market share model before liberalization with all airline-related variables

	before liberalization		
	LIS-PDL-LIS	LIS-PDL	PDL-LIS
$price_{day}$		0.15	
$price_{2m}$			
$freq_{day}$	3.49	3.09	4.15
$freq_{week}$		1.92	
$time\_points$			
_const	-1.34	-2.81	-1.96

This section is divided into several sub-sections which are oriented on the market segment. In Section 6.1.4, route LIS-PDL-LIS is looked at as a return flight. This means that the the values for the attributes for both flights are brought together. In the following sections, the results for the the flights from Lisbon to Ponta Delgada and back from Ponta Delgada to Lisbon are analyzed in Section 6.1.4 and 6.1.4 respectively.

LIS-PDL-LIS

In this section the flight between Lisbon, Ponta Delgada and and back to Lisbon is considered. The results of the GLM calculation with all the airline-related variables can be found in the corresponding column in Table 6.4. The independent variable on_time is also not included in the model because of problems in the validation.

As can be seen in Table 6.4, the only independent variable that is significant in this model is the $freq_{day}$. The variable $freq_{day}$ is actually the frequency share of SATA compared to the competitors. From the flight details it is known that TAP flew with an A319 whereas SATA used its A320 and was able to offer business seats. Therefore, the frequency share also gives an indication about the passenger share the airlines are able to transport. The passenger share is another name for the market share because it is calculated by dividing the passengers of SATA over the total number of passengers transported on that route on that day. Therefore, it makes sense that $freq_{day}$ is significant in the model.

Furthermore, a constant is added to the model which is significant. It is negative which means that if all the variables are set to zero, SATA is unable to achieve a market share of above 50%. The remaining influence is triggered by $freq_{day}$.

The results of the model show that neither of ticket price attributes are significant. The market share before liberalization is in favor for the Azorean airline SATA. It lies on average for the LIS-PDL-LIS route at 77% for SATA and consequently 23% for TAP. The average ticket price for the route when booking SATA is about 45€ higher than for TAP even though the ticket price is higher. No clear evidence can be drawn that one of the ticket price independent variables had influence on market share before liberalization.

LIS-PDL

The results of the market share model for the time before liberalization and for the flights from Lisbon to Ponta Delgada are shown in Table 6.4. So only a one way direction is considered. It helps to understand the results of the combined market share model when looking at the market sections separately.

The results of this case show that the independent variable $freq_{day}$ and the constant are significant, that are also found in the previous full return case. However in comparison to the previous case, also $price_{day}$ and $freq_{week}$ are significant. The coefficient that represents the $price_{day}$ is significant and positive which is not as expected.

The reason for this can be found in the attribute input data. The values of $price_{day}$ increase with increasing market share which is a phenomenon that is not as expected. Now, it is important to analyze if the raw data of SATA or TAP is the main influence on this problem. When plotting the raw ticket prices the day prior departure over the market share for both airlines it becomes clear that for SATA the ticket prices are relatively constant but decrease slightly with increasing market share. For TAP however, the ticket price the day prior departure increases with increasing market share. The independent variable $price_{day}$ is calculated by dividing the ticket price of SATA over the average ticket price of the competitors. Therefore, it can be concluded that the ticket price the day prior departure of TAP is the reason why the β-coefficient of $price_{day}$ behaves different as expected.

Another independent variable that was not significant in the full return case before liberalization is $freq_{week}$. In the data it can be seen that the market share increases with increasing $freq_{week}$ value. SATA increases their frequency per week towards the summer months because of increasing demand. TAP keeps their frequencies steady at four frequencies per week. Market share increases towards the summer and the value for $freq_{week}$ also increases because TAP does not change its number of frequencies. Therefore, it is possible to find a significant relationship between the two variables.

PDL-LIS

This section has the same intention than the previous, which is to understand the results of the LIS-PDL-LIS case for the situation before liberalization. In this section the results for the flights departing from Ponta Delgada and arriving in Lisbon are analyzed. The β-coefficients can be found in the last column in Table 6.4

The results of the market LIS-PDL before liberalization fully support the results of the full return flight from Lisbon to Ponta Delgada and back. In this model $freq_{day}$ is also the only significant independent variable. The value for the β-coefficient of $freq_{day}$ is higher than in the full return model because on average the value for $freq_{day}$ is slightly lower in this case compared to the LIS-PDL-LIS case whereas comparable average market shares are achieved.

6.1.5. Results of the market share model before liberalization without $freq_{day}$ as independent variable

In this section the results of the market share model for the situation before liberalization are presented. The difference to the previous section is, that in this case, the independent variable $freq_{day}$ is disregarded from the model. The significant β-coefficients for the different markets in this case can be found in Table 6.5.

The reason for creating a market share model that disregards $freq_{day}$ lies in the fact, that $freq_{day}$ is prone to be highly correlated with the market share as already explained in Section 6.1.2. If an independent variable is highly correlated to the dependent variable, it dominates the contribution to market share and "pushes" other independent variables out of the model which would be significant if $freq_{day}$ would not be present in the market. Therefore, the significance of other independent variables apart from $freq_{day}$ are determined and analyzed in this section.

Table 6.5: Results of the market share model before liberalization without $freq_{day}$ as independent variable

	before liberalization		
	LIS-PDL-LIS	LIS-PDL	PDL-LIS
$price_{day}$		0.15	
$price_{2m}$	-0.05		
$freq_{week}$	5.43	3.88	11.26
$time\_points$			0.42
on_time			
_const	-3.20	-1.99	-8.74

LIS-PDL-LIS

The first market that is analyzed is the full return market between Ponta Delgada and Lisbon for the situation before liberalization without taking the independent variable $ferq_{day}$ under consideration. The β-coefficients of the significant variables in the model can be found in Table 6.5. It can be seen that $price_{2m}$ and $freq_{week}$ are the significant variables.

Additionally, a constant is added to the model. The constant contributes for 4 percentage points to the market share which is significantly lower that the constant contributions found in the previous cases. The reason for this might be that the airlines that are competing against each other on this market before liberalization are SATA and TAP. Those airlines are both Portuguese airlines and both legacy carriers. The constant includes all influences of factors that are not taken into account in the independent variables. The remaining influence could be difference in service or loyalty of the costumer. However, the difference between the airlines in the market is little which is represented by the low contribution value of the constant.

The β-coefficient of the variable $price_{2m}$ is negative which is expected. The market share is expected to decrease when the ticket prices will increase. The contribution of $freq_{week}$ on market share behaves positively. Market share increases when the value for $freq_{week}$ increases. The negative and positive behavior on market share can also be found in the input data of $price_{2m}$ and $freq_{week}$ respectively when plotting the data over the market share.

LIS-PDL

The full return market is split into its two segments. First the route between Lisbon and Ponta Delgada is analyzed in this section. The resulting β-coefficient of the significant variables in the model when disregarding the independent variable $freq_{day}$ can be found in Table 6.5. It can be seen in the table that the significant variables are $price_{day}$ and $freq_{week}$. Also a significant constant is present in the model.

The sign of the ticket price variable $price_{day}$ is positive which is not as expected. When the ticket price increases, the market share is suppose to decrease. When $price_{day}$ is plotted against the market share one can see the slightly increasing trend in $price_{day}$ when increasing the market share. It is of interest to find out if this is cased by the ticket prices of SATA or of the competitor TAP which are the only two airlines that are flew in that market. This can be found out by plotting ticket price the day prior departure of SATA and TAP as well as the market share of SATA over the time period. Whereas the trend line of the ticket price of SATA is relatively steady around 100€, the ticket price of TAP increases towards the summer months. Also the market share of SATA increases when coming closer to the summer. This effect results in a positive β-coefficient of $price_{day}$.

The remaining significant variable is $freq_{week}$. The sign of the coefficient is positive which is expected since the market share is supposed to increase when the frequency per week is increased. Therefore, the variable is properly represented in the market share model.

These two variables are also significant in the case when the independent variable $freq_{day}$ is not disregarded from the model. They show a strong relationship with market share since they are not "pushed out" of the model by the correlated variable $freq_{day}$.

PDL-LIS

The final market in this section is the market between Ponta Delgada and Lisbon. The results of the model for this market for the situation before liberalization without considering $freq_{day}$ as an independent variable, can be found in Table 6.5. The significant variables in this model are $freq_{week}$ and $time\_points$. A constant is added to the model too. The value of the constant and the β-coefficient of $freq_{week}$ are significantly higher than the values for the constant and coefficients seen in the previous models.

When plotting the $time\_points$ input data over the market share, a positive relationship can be found. The market share increases when the $time\_points$ value increases. A positive relationship can also be found in the input data for the independent variable $freq_{week}$ and the market share.

The unexpected high values however, raise doubt if the model is suitable to represent the market share in the market between Ponta Delgada and Lisbon before liberalization. The input data counts only 65 observations which might cause problems in the calibration. Also the input values of both significant variables do not have much variation. The difference of the maximum and minimum value of the variable $freq_{week}$ is only 6.7 percentage points and the value for $time\_points$ takes over either 0.5 or 1. Slight changes in the variable will tremendously change the market share for that day. Therefore, the model is prone to give faulty results.

6.2. Impact

In this section the impact which each variable has on market share is analyzed. In the previous sections, the β-coefficient are shown for the different markets under investigation. However, no straight interpretation about the impact can follow from the β-coefficient directly, because it is dependent on the magnitude of the independent variable. Therefore, a new parameter $impact_j$ is introduced and described in more detail in Section 4.7. This value is a percentage that indicated how much each variable contributes to the market share. This way it is possible to compare variables between each other, different routes and even different situations like the situation before and after liberalization.

Therefore, the results of the impact calculation is given in the following. It is divided into the different market share models that are created with different independent variables. First the impact of the variables in the market share model with all airline-related variables is presented in Table 6.6.Then the impact of the same model is reduced by the independent variable $freq_{day}$ is shown in Table 6.7. Finally, the first model is expanded by monthly dummy variables and the impact of this model can be seen in Table 6.8. In the tables the impact of each variable on market share can be found and in the following each of these variables are discussed separately.

Table 6.6: Impact of the variables in the market with all airline-related variables

	after liberalization			before liberalization
	LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO	LIS-PDL-LIS
$price_{day}$		9.75%		
$freq_{day}$	48.12%	47.15%	78.64%	37.88%
$freq_{week}$			21.36%	
_const	51.88%	43.10%		62.12%

Table 6.7: Impact of the variables in the market with the airline-related variables disregarding $freq_{day}$

	after liberalization			before liberalization
	LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO	LIS-PDL-LIS
$price_{day}$		36.71%		
$price_{2m}$	35.16%			7.71%
$freq_{week}$			21.84%	12.66%
$time\_points$	28.02%	35.93%		
_const	36.82%	27.36%	78.16%	79.63%

Table 6.8: Impact of the variables in the market with monthly dummy variables

	after liberalization		
	LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO
$price_{day}$			5.81%
$freq_{day}$	48.12%	40.79%	30.35%
may		3.14%	
june		4.68%	7.58%
july		7.45%	11.74%
_const	51.88%	43.95%	44.51%

6.2.1. Impact of ticket price on market share

In this section the impact of ticket price on market share is analyzed. In the market share model two different ticket price variables are introduced as independent variables which are ratio of the ticket price the day prior departure ($price_{day}$) and the ratio of the ticket price two months prior departure ($price_{2m}$). The impact of these variables on market share are discussed separately in Section 6.2.1 and 6.2.1 respectively.

Impact of the independent variable $price_{day}$

The independent variable $price_{day}$ represents the ratio of the ticket price the day prior departure of the airline SATA compared to the average ticket price of the competitors. It can be seen in the tables that the variable is significant in the market LIS-PDL-LIS in the market share model with all airline-related variables and in the model where $freq_{day}$ is disregarded. The importance of this variable increases from 9.75% to 36.71% when $freq_{day}$ is disregarded. This makes sense because $freq_{day}$ is highly correlated with market share and therefore, the impact is higher too. Another market where $price_{day}$ is significant is the market between Porto and Ponta Delgada when the monthly dummy variables are introduced to the market share model. It takes over a small share of 5.81% of the total impact.

The independent variable $price_{day}$ has impact on market share in both full return markets after liberalization but appears in different market share models that have different independent variables other that $price_{day}$. Booking a flight ticket the day before departure seems to be common in the market between the Portuguese mainland and the Azores.

Impact of the independent variable $price_{2m}$

The impact of the independent variable $price_{2m}$ in market share can only be detected in the combined return market LIS-OPO before liberalization and in the market between Lisbon and Ponta Delgada before liberalization where the independent variable $freq_{day}$ is deleted from the model. In the market LIS-OPO the variable takes over 35.16% of the total impact on market share. In the market LIS-PDL-LIS after liberalization it only takes over 7.71% of the impact.

The independent variable $price_{2m}$ is only significant when $freq_{day}$ is deleted from the model. Therefore, there needs to be a relationship between $price_{2m}$ and $freq_{day}$. $Price_{2m}$ takes over parts of the impact that was triggered by $freq_{day}$ in the previous market share model. This relationship is know from basic economics which is called price elasticity of demand. The demand is dependent on the price and the quantity of a product. The quantity can be translated to the seats offered on this route which is directly related to the flight frequency. This relationship is the reason why $price_{2m}$ only appears in the market share model where $freq_{day}$ is disregarded.

Concluding remarks about the impact of ticket price on market share

In literature it is found that ticket price is the driving factor behind the passenger choice between different airlines. However, in the market share model this is only the case when the frequency variables are not significant. The impact of the ticket price variables is significantly smaller than the impact of the frequency variables if both are present in the model. Therefore, a relationship between frequency and ticket price is expected. The relationship is called price elasticity of demand.

In the data set that is created to analyze the market share of the Azorean case study, the variable $freq_{day}$ is highly correlated and therefore, takes over parts of the impact that would be allocated to the ticket price variables otherwise. In the tables it can be seen that when $freq_{day}$ is deleted from the market share model, one of the ticket price variables appear in the model.

Furthermore, it should be noted that there is a difference between perceived ticket price and actual ticket price. It is expected that ticket price takes over a major contribution to the influence on market share which is based on the feeling that LCC's always offer significantly lower prices than legacy carriers. The model however is based on the ratio of the actual ticket price. For the LCC's, the ticket price will increase when the booking date comes closer and closer to the departure date. Therefore, the ratio of $price_{day}$ is generally lower than $price_{2m}$.

6.2.2. Impact of frequency on market share

In this section the impact of frequency on market share is analyzed. In the market share model, two different independent variables that represent frequency can be found. These two variables are $freq_{day}$ and $freq_{week}$. The impact of $freq_{day}$ and $freq_{week}$ is discussed in the following.

Impact of the independent variable $freq_{day}$

At first glance it can be seen, that the $freq_{day}$ is the variable that contributes the the market share in all situations. An increase in $freq_{day}$ results in the fact that passengers are able to choose between more flights and more seats are available on the route. $Freq_{day}$ is directly linked to market share which means in other words they are highly correlated. This is the reason why it is the only or one of two significant independent variables in the model and takes over a major part of the impact on market share. When all airline-related variables are implemented in the model, the impact of $freq_{day}$ increases from 37.88% to 47.15% due to liberalization in the market LIS-PDL-LIS. This mean that the airline attribute frequency per day is more important in the situation after liberalization than before liberalization.

In the market share model with monthly dummy variables, $freq_{day}$ is the strongest source of impact that can be assigned to one independent variable directly. Only the constant has more impact on market share. The impact equals to 48% for the combined market LIS-OPO where no monthly dummy variables are significant. In the two separate markets, monthly dummy variables are found to be significant which decreased the impact value. However, it is still as high as 40.79% and 30.35% for the markets LIS-PDL-LIS and OPO-PDL-OPO respectively.

In the market OPO-PDL-OPO with all airline-related variables, $freq_{day}$ takes over an unusual high share of the impact. The reason for this is the low amount of observations that are available for this model which makes the constant not to be significant and $freq_{day}$ taking over the share of the impact that would be appointed to the constant otherwise.

Impact of the independent variable $freq_{week}$

The independent variable $freq_{week}$ has only impact on market share in the models without monthly dummy variables and on the route OPO-PDL-OPO after liberalization as well as on the route LIS-PDL-LIS before liberalization. It stands out that the impact of $freq_{week}$ barely changes when $freq_{day}$ is included in the model or disregarded. For the market LIS-PDL-LIS before liberalization, $freq_{week}$ takes over parts of the impact of $freq_{day}$ when this variables is deleted. Since the input values for the frequency per day are the summed values of the frequency per day of each airline, these two variables are linked. This linkage explains that $freq_{week}$ becomes significant when $freq_{day}$ is deleted from the model.

The variable $freq_{week}$ represents flexibility and continuity for the passenger. This is especially interesting when the airlines do not offer daily flights. Passenger make their plans at which day they would like to travel and if an airline does not offer a flight on that particular day, the passenger is likely to switch to another airline. This explains why $freq_{week}$ is only significant in two markets: OPO-PDL-OPO after liberalization and LIS-PDL-LIS before liberalization. Even though SATA offers daily flights in these markets (with some exceptions), SATA's competition does not offer daily flights. This effects the value of $freq_{week}$ since it is the frequency share of SATA compared to its competitors. This means that the passengers are more likely to fly with SATA if the competing airline does not offer continuous frequency of flights during the week.

Concluding remarks about the impact of frequency on market share

The independent variables that represent the frequency have the highest impact on market share in all the models. Only the constant influences the market share more. The impact of $freq_{day}$ increases due to liberalization. $Freq_{week}$ takes over shares of the impact of $freq_{day}$ when this variables is deleted from the model due to correlation.

6.2.3. Impact of the independent variable $time\_points$ on market share

The independent variable $time\_points$ represents the frequency share of the flights that depart in the time window between 7-9 o'clock in the morning and between 17-19 o'clock in the afternoon. These time windows are found to be preferred for the passengers.

The impact of $time\_points$ can only be found in the market share model where $freq_{day}$ is deleted from the model. Therefore, a relationship between $time\_points$ and $freq_{day}$ needs to be present. This makes sense because if $freq_{day}$ increases, it is also more likely that the extra flight falls in one of the time windows and consequently increases the value of $time\_points$. Therefore, $time\_points$ takes over a part of the impact that was assigned to $freq_{day}$. Since it is connected to $freq_{day}$ the impact of $time\_points$ is rather high when significant. In the market LIS-PDL-LIS it has an impact of 35.93% whereas the impact of $time\_points$ in the combined market LIS-OPO equals to 28.02%.

6.2.4. Impact of the independent variable on_time on market share

The independent variable on_time did not appear in either of the market share models. The reason for this is threefold. In the model with all airline-related variables, the variables was not included because with the variable, it was not possible to validate the model. In the market share model where $freq_{day}$ was disregarded, on_time was not found to be significant. And finally in the model with the monthly dummy variables, the variable was highly correlated with at least one of the monthly dummy variables. Therefore, the variable was also excluded from the market share model with monthly dummy variables.

Therefore, it is possible to conclude, that the on-time performance of an airline has not direct impact on market share. SATA is known for its poor on-time performance but this seems to not influence the passenger airline choice and therefore, the market share. However, no accurate data could be found for the on-time performance of the LCC's. The LCC's only publish performance data for their full market. This data is used and assumed to be suitable for the market between the Azorean archipelago and Portugal mainland. There is a possibility, that the actual effect of the on-time performance is included in the impact that constant has on market share because the constant includes all the effects that are not specifically expressed in one of the independent variables.

6.2.5. Impact of the monthly dummy variables on market share

In Table 6.8 the impact of the independent variables on market share can be seen for the market share model that includes monthly dummy variables. The variables in introduced to indicate the impact of time after liberalization and the fact that SATA sold seats on their flights even before the entrance of the LCC's was announced. It can be seen from the table that in the market LIS-OPO none of the monthly dummy variables can be identified to have impact on the market share. In the split markets however, the impact of time can clearly be seen.

In the market LIS-PDL-LIS, the impact of the monthly dummy variables increases from May to July. The dummy variables compare the month April to the corresponding month. This means that compared to April, flying in May has 3.14% more impact on market share whereas flying in July even has 7.45% more impact on market share. The reason for this lies in the seasonality of the Azores. The demand is higher in the summer months because many leisure tourists spend their vacation on one of the islands. Also, the adaption of the liberalization can be slower in peoples minds. This means that the LCC's need to warm up in the market until passengers will choose to fly with them.

In the OPO-PDL-OPO market the impact on market share also increases in June with 7.58% and in July with 11.74%. However, when looking at the β-coefficient of this model in Table 6.3 it can be seen that if a flight departs in June or July, it actually has negative effect on the market share compared to the month April. The reason for this lies in the fact that SATA sold tickets for the flights after April even before it was announced that the LCC's will enter the market. These tickets were purchased directly via SATA or sold to travel agencies. The average market share of SATA in April was 46.8% and this value decreased to 39.6% in July which means that Ryanair was able to achieve an increasing amount of market share. It also appears that Ryanair needed to warm up a little in the market before taking over a severe share of the market from the SATA that had a monopoly position before liberalization.

6.2.6. Impact of the constant on market share

The constant is present in all market share models except the model for the OPO-PDL-OPO market with all airline-related variables. It takes over a significant share of the impact ranging from 27.36% in the LIS-PDL-LIS market before liberalization to 79.63% in the LIS-PDL-LIS market after liberalization (both for case when $freq_{day}$ is disregarded).

When comparing the market LIS-PDL-LIS before and after liberalization in Tables 6.6 and 6.7, it can be seen that the impact of the constant is higher for the situation before liberalization for both market share cases. This means that the importance of the effect that are not included in any of the independent variables was higher before liberalization. Those effects could include certain services like extra baggage price, food on-board or in-check services, or loyalty of the passengers. This makes sense because before liberalization, only legacy carrier competed on this route which means that the difference between TAP and SATA is not as severe as the difference between those airlines and LCC's. Therefore, little differences like the loyalty or the difference in food can make significant impact on the passenger choice behavior and consequently the market share.

Furthermore, for the situation after liberalization, the impact of the constant decreases when the independent variable $freq_{day}$ is deleted from the model. Then, the impact is more spread over the remaining significant variables and the constant. For the situation before liberalization the opposite is the case.
The impact of the independent variables for the market share model where monthly dummy variables are added to the model is comparable to the case with all airline-related variables for the markets LIS-OPO and LIS-PDL-OPO as can be seen in Table 6.8. However, in OPO-PDL-OPO the impact decreases because parts of the impact are taken over by the significant monthly dummy variables.

6.3. Deleted observations

It needs to be pointed out that if there is a value missing for one of the variables, the entire data point is deleted from the calculation. A missing value can be caused by monopoly position of one airline on that day or if a flight is sold out on one day no ticket price can be evaluated. This methodology of removing observations is called list-wise deletion.

If a market share value of 0 or 1 is present on a particular day, the full observation is deleted from the model. The reason behind the deletion is that the model would take into account effects that are not actually present. This means that for example if the ticket price could be raised tremendously or the frequency is increased to a level that it would be way beyond the demand, it would not have any effect on the market share because it is bound to either 0 or 1 since no competition is present.

This happens regularly in the market OPO-PDL-OPO after liberalization and LIS-PDL-LIS before liberalization because not all airlines in the market offer daily flights. Ryainair offers daily flights throughout the week except for Saturday between Ponta Delgada and Porto. SATA stepwise increased its flights from 5 return flights per week to daily flights. This leads to 3 day of monopoly position of one of the airlines in April to just one day in July. In the market between Ponta Delgada and Lisbon the airlines have steady schedules that include daily flights from SATA and TAP offers 4 return flights per week.

Furthermore, excluding observation where no ticket price could be determined because the flight was sold out has significant effect on the results because the time span includes July which is a traditional peak month for the Azorean market. Especially, when collecting the data for the LCC's it regularly occurred, that the ticket price the day prior departure could not be determined because the flight was sold out. This leads to undersampling of the months with high demand.

Both effects reduce the number of observations from 123 to 57 observation in the OPO-PDL-PDL market and from 123 to 72 observation in the market LIS-PDL-LIS before liberalization. This makes the model less stable to changes and results in unexpected behavior as detailedly explained for each model in Section 6.1.

Is it possible to run the model with the observations that represent monopoly position of one of the airlines. The results can be found in Table .... It can be seen that for the markets LIS-OPO and LIS-PDL-LIS after liberalization, including the observation has limited effect on the β-coefficients. For the remaining market however, the results change.

In Table 6.9 the results can be found of the GLM calculation when the observations are not deleted on days, where monopoly position of one of the airlines is present. It can be seen that for the market OPO-PDL-OPO the value of $freq_{day}$ is unusually high and for $freq_{week}$ the value for the β-coefficient is unusually low. Also, the constant is higher than 0 which results in a contribution of more than 50 percentage points. In the results for LIS-PDL-LIS before liberalization the value for $freq_{day}$ is also high and the coefficient of $price_{2m}$ has a positive sign where a negative sign would be expected. Both models do not coincide with the results shown in the previous section.

To check if the market share models are suitable, the validation is performed. Both market share models could not be validated. For the market OPO-PDL-OPO the calculated market share for the reference month September is so small that it is negligible. The market LIS-PDL-LIS before liberalization is not possible to validate because no validation data is available.

Table 6.9: Results of the market share model including monopoly observations

	after liberalization	before liberalization
	OPO-PDL-OPO	LIS-PDL-LIS
$price_{day}$		-0.15
$price_{2m}$		0.065
$freq_{day}$	11.62	7.61
$freq_{week}$	-9.72	
_const	0.95	-4.09

6.4. Fit of the model

In Section 6.1, the results of 3 different market share models are presented. These models are all market share models but differ in the independent variables that are used. The first model that is indicated with "all" in Table 6.10 indicates the market share model with all airline-related independent variables. The second model "without $freq_{day}$" is similar to the previous model but the independent variable $freq_{day}$ is not considered. The third model, expands the first model by "monthly dummies". In Table 6.10, the values for the pseudo R^2, Deviance and the Pearson's goodness of fit can be found for all models in all return markets for before and after liberalization.

There is a problem with these goodness of fit parameters. All of them are able to indicate which model fits better relative to each other but not absolute [23] [45]. This means that only model can be compared which use the same input data and have the same output. Therefore, the three market share models can be compared for the same market but no conclusion can be drawn if the market share model is a good fit for the input data. Also, it is not possible to compare the different markets because the input data is not the same.

Table 6.10: Results of the model fit calculation

		after liberalization			before liberalization
		LIS-OPO	LIS-PDL-LIS		LIS-PDL-LIS
pseudo R^2	all	0.02	0.031	0.0046	0.0053
	without $freq_{day}$	0.0048	0.012	0.0018	0.0027
	monthly dummies	0.02	0.033	0.0087	
Deviance	all	0.53	1.92	0.63	1.02
	without $freq_{day}$	1.81	4.45	0.84	1.22
	monthly dummies	0.53	1.56	0.35	
Pearson	all	0.53	1.9	0.63	0.96
	without $freq_{day}$	1.75	4.31	0.84	1.16
	monthly dummies	0.53	1.55	0.35	

All 4 indicators of the goodness of fit have a tendency when a model fits the data better. For the pseudo R^2 a high value is a signal of a good fit. The value for the R^2 are bound between 0 and 1 whereas the pseudo R^2 lies also within these boundaries but usually does not reach as high values as the R^2 [45]. For the Deviance and the Pearson's goodness of fit a low value indicates a better fit when comparing two models.

As can be seen in the table, the market share model with monthly dummies fits the data best. It scores the highest pseudo R^2 values and the lowest values for deviance and for the Pearson's goodness of fit. Two markets behave differently. In market LIS-OPO, the results for the market share model with all airline-related variables and the market share model with monthly dummy variables have only $freq_{day}$ as significant independent variable. The results are identical and therefore, the values for the fit of the model is identical too. Furthermore, no market share model with monthly dummy variables was created for the market LIS-PDL-LIS before liberalization. This is the reason why no value can be found in the table for this market share model.

It makes sense that the market share with the monthly dummy variables fits the data best because it is an extension to the market share model with all airline-related variables. The independent variable that is always significant is $freq_{day}$. Often, it is possible to create a significant model with this variable only. Therefore, it can be expected that the model without $freq_{day}$ fits the data the least. The market share model with monthly dummy variables give more options in the independent variables and therefore, a better model can be created.

6.5. Summary of the results

In this research, 3 different market share models are created to draw conclusion about the impact of liberalization on market share: one with all airline-related independent variables, the second where the variable $freq_{day}$ is disregarded and the last that includes all airline-related independent variables and monthly dummy variables.

The variables $freq_{day}$ is the most important independent variables because it is significant in all market share models (except when disregarded) and takes over a significant share of impact on market share. This impact varies between 30-78%. The impact of $freq_{day}$ increases due to liberalization which means that it is more important to have a higher frequency than the competition after liberalization. Furthermore it can be said that the impact of $freq_{day}$ decreases when monthly dummy variables are added to the model.

The constant is another source of impact on market share that is significant in almost all market share models. It represents all effects that are not taken into account in the independent variables. Is exceeds the impact value of all independent variables in all but two models. The impact ranges from 27% to almost 80%. The constant was more important before liberalization than after liberalization. Also, the impact of the constant decreases when $freq_{day}$ is removed from the model for the situation after liberalization - for the situation before liberalization, the impact increases.

The ticket price variables are primarily significant in the market share model where $freq_{day}$ is disregarded. Additionally, $price_{day}$ has impact on market share for all airline-related variables in the LIS-PDL-LIS market after liberalization. The variables $freq_{week}$ and $time\_points$ also take over a big part of the impact when $freq_{day}$ is disregarded. The independent variable on_time is not significant in any of the market share models.

Finally, the monthly independent variables show increasing impact towards the summer months in the market LIS-PDL-LIS after liberalization. On the route between Porto and Ponta Delgada the impact is also increasing towards the summer the β-coefficients are negative which means that the market share decreases towards the summer. This could have been triggered by the fact that seats of SATA flights were sold before the liberalization was announced.

7

Sensitivity Analysis, Verification and Validation

To be able to say something about the quality of the developed market share model, the model has to be tested. This is done in three steps. First a sensitivity analysis is performed to find out how the market share reacts if there are changes in one of the attributes. The results of this analysis can be found in Section 7.1. In Section 7.2 the market share model is verified. Finally in Section 7.3, the validation is shown where the market share that is prospected by the model is tested against the actual market share values.

7.1. Sensitivity Analysis

In the sensitivity analysis the change of impact of one attribute on the dependent variable is tested. To do this the independent variable under consideration is changed to the minimum and maximum value. All other independent variables are kept constant at the mean value. In Table 7.1, the percentage change in market share can be found that is triggered by setting the independent variable to the minimum and maximum value. Note, that if setting an independent variable, that has a negative β-coefficient, to its maximum value, the change in market share will also be negative. In Table 7.1 it can be seen that this is the case for the independent variables $price_{day}$ and, against expectations, $freq_{week}$.

As can be seen in Table 7.1, changing the independent variable $freq_{day}$ to its minimum and maximum value, the change in market share is highest. In the combined market LIS-OPO, the change in market share is -22.58% when setting to its negative value and 18.1% when setting to the maximum value. In the LIS-PDL-LIS market after liberalization, the change for the minimum value is lower with 20.95% but it achieves a market share change of 24.3% when $freq_{day}$ takes over the minimum value. The market share changes in the OPO-PDL-OPO market triggered by $freq_{day}$ are lower, and in the situation before liberalization in the market LIS-PDL-LIS, the changes triggered by $freq_{day}$ are the lowest compared to the remaining markets.

Table 7.1: Results of the sensitivity analysis

		after liberalization			before liberalization
		LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO	LIS-PDL-LIS
price_day	min		2.40%		
	max		-7.42%		
freq_day	min	-22.58%	-20.95%	-13.12%	-8.76%
	max	18.10%	24.30%	12.24%	5.96%
freq_week	min			3.48%	
	max			-3.55%	

In the LIS-PDL-LIS market after liberalization, $price_{day}$ is also significant and a change of 2.4% and -7.42% is achieved when the independent variable is set to its minimum and maximum value respectively. Also in the market between Ponta Delgada and Porto after liberalization, there is another variable significant next to $freq_{day}$ which is $freq_{week}$. This variable has a unexpected sign of its β-coefficient which is discussed in Section 6.1.1. The change in market share in this case equal to 3.48% when the value for $freq_{week}$ is the minimum value, and -3.55% when it is the maximum value.

This sensitivity analysis support the findings that were made in Section 6. The impact of $freq_{day}$ on market share is higher after liberalization than before liberalization. Furthermore, $freq_{day}$ is the independent variable that is most important (next to the constant). The variable triggers the highest changes in market share. Additionally, market share is less sensitive to changes of the remaining variables.

7.2. Verification

In the verification it is tested if the market share model is developed in the right way. This means that there are certain requirements the model needed to have to be suited for the project and in this section one needs to check if these were met.

In this case, the verification is the discussion if the signs of the coefficients are as expected which is already evaluated in Section 6 where the results are shown. In the full return market between Ponta Delgada and Lisbon all β-coefficients have the expected sign. This cannot be said about the route from Porto to Ponta Delgada and back. In this route in the market share model with all airline related variables and the market share model without the independent variable $freq_{day}$, the significant coefficient of $freq_{week}$ has a negative sign instead of a positive. The reasons for that are broad and explained in detail in Section 6.

7.3. Validation

In the validation, the model is tested if it represents the reality properly. The model that is developed in this thesis is a market share model. The Azorean archipelago is taken as a case study because the model is supposed to be implemented in a newly liberalized low-demand market. In the next step, the results of the validation is shown that was carried out with the passenger input data of September 2015. It is tested whether the market share, that is calculated the same manner as presented in Section 5.2.1, is comparable to the market share that is generated using the β-coefficients for each market (explained in more detail in Section 6). To differentiate these two market shares, the terms **actual market share** and **prospected market share** are introduced.

The actual market share represents the market share that is based on taking the airport report of September 2015, approximating the passenger numbers of each day from the figure that is presented in the report and subtract all the known passenger numbers that do not belong to the market that is being analyzed. For the market LIS-PDL-LIS after liberalization, the passenger numbers of TAP and SATA are known on a daily basis but the passenger numbers of the Ryanair on the route OPO-PDL-OPO and the passenger numbers of the charter airlines are only known per month. Therefore, the passenger numbers that are only known per month need to be spread over the full month. The charter airlines offer weekly frequencies which results in the fact that the passenger numbers are spread evenly over the month. Ryanair offers 6 frequencies per week from Ponta Delgada to Porto which means that weekly seasonality is taken into account when spreading it over the month. With this method the actual passenger data is used to determine the actual market share.

The prospected market share is based on Equation 7.1 which is already presented in another form in Section 4.1. The β-coefficients that are determined for each market are presented in Section 6. For the validation, the significant β-coefficients are multiplied with the value of their according attribute and are then summed. This is done for each observation, e.g. for each airline each day of the month September 2015. By doing so, a prospected market share can be calculated for each observation. This prospected market share can then be compared with the calculated market share.

$$marketshare = \frac{exp(\beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots)}{1 + exp(\beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots)} \tag{7.1}$$

The validation is performed for the combined and full return routes between Ponta Delgada and Lisbon as well as Porto for the situation after liberalization. There 3 markets are validated for 2 market share models, one with all airline related variables and one without $freq_{day}$ as independent variable. The market share model with monthly dummy variables is not validated since the monthly dummies only span the time window of April to July and not September. Validation would only be possible if validation data is available for the months the monthly dummies are present. The results of the validation can be found in the following sections sorted by markets.

For the situation before liberalization there is no actual validation data available. When prospecting the market share for the same time period as the input data, the market share model performs very well which is expected. However, this is no validation for the market share model and the model can only be validated with the according validation data.

7.3.1. Results of the validation of the market LIS-OPO after liberalization

The market share model of the combined market between Ponta Delgada and Lisbon as well as Porto needs to be validated. The results of the validation calculation can be found in Table 7.2. The market share model with all airline related variables achieves a prospected market share of 39.19% which is close to the actual market share of 40.1%. The actual and prospected market share of this maket is

Table 7.2: Results of the validation after liberalization

		LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO
all	Actual	40.10%	37.45%	45.85%
	Prospected	39.19%	37.04%	58.32%
without freq_day	Actual	40.10%	37.45%	45.85%
	Prospected	32.58%	34.65%	43.46%

visualized in Figure 7.1. It can be seen that the two lines are comparable and follow the same trend. Therefore, this market share model is validated.

The market share model where $freq_{day}$ is deleted from the model doe snot perform as good in the validation as the previous model. The results of the fit of the model in Section 6.4 already indicated that the model with all airline related variables fits the available data better than the model without $freq_{day}$. This can also be seen in the validation. The actual market share market share stays the same but the prospected market share of this model is equal to 32.58%. Therefore, the difference is 7.52%. This means that the market share model with all airline related independent variables is preferred for the combined market.

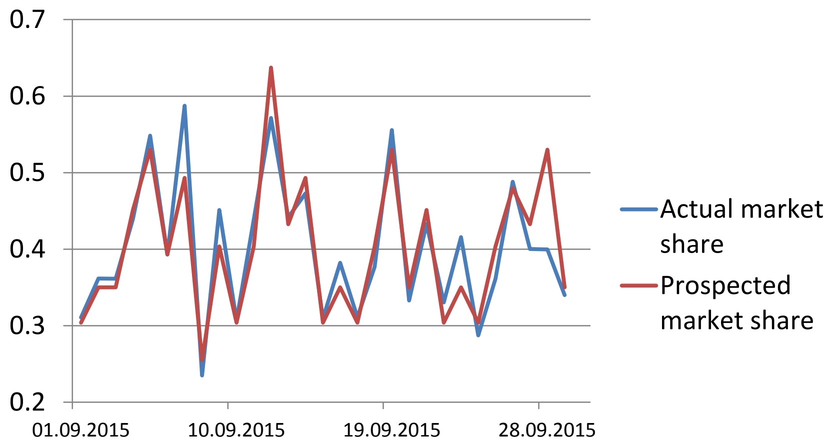

Figure 7.1: Actual market share vs. prospected market share in the combined market after liberalization for the market share model with all airline related variables

To conclude the validation of the market share model for the combined for the situation after liberalization, it can be said that the model works well for forecasting the market share per month. Therefore, the model should be used when the user wants to calculate the market share per month and not too much emphasis should be laid on the prospected market share per day. As can be seen in Figure 7.1, the actual and prospected market share follow the same trend but are not approximately equal when looking on a daily basis.

7.3.2. Results of the validation of the market LIS-PDL-LIS after liberalization

The results of the validation with the data between Ponta Delgada and Lisbon after liberalization can be found in Table 7.2. The actual market share for this route is equal to 37.45% for September 2015. The prospected market share is calculated using the method described above. For the market share model with all airline related variables, the prospected market share achieves a value of 37.04% which means that its difference is only 0.11%. These values are very close to each other which is the reason why the model is concluded to be validated. The market share model seems to be suitable to prospect market shares in the newly liberalized Azorean market.

For the market share model, in which the independent variable $freq_{day}$ is not considered, the actual and prospected market share further apart from each other than in the previous model. The difference is equal to 2.8% since the prospected market share is 34.65%. This difference is still considered to be satisfactory to validate this market share model. Similar things count for this market than for the combined market, the market share models are supposed to be used to forecast market share per month because if it is used per day, major differences between actual and prospected market share may occur.

7.3.3. Results of the validation of the market OPO-PDL-OPO after liberalization

In Table 7.2 the validation results of the market between Ponta Delgada and Porto for the situation after liberalization can be seen. The actual market share for this market is equal to 45.85% in September 2015. For the market share model with all airline related variables as independent variables, the prospected market share is 58.32%. This is far off from the actual market share and not sufficient to validate the model. The data available is the main problem with the validation of the market share model for this particular market. The data is limited which is due to the fact that there are several days where one of the airlines have monopoly position. Therefore these observations are deleted from the model and cannot be used for the validation. Since the validation data is only one month, deleting the observations of several days have impact on the validation.

The average prospected market share for the market share model where $freq_{day}$ is deleted from the model, is found to be 43.46%. This is close to the actual market share. It is surprising that this model performs better without $freq_{day}$ and having only one significant variables which is $freq_{week}$. The reason for this is, that $freq_{week}$ equals to 0.5 on most of the days in September. For this value specifically, the market share would be 43.78% which is close to the actual market share. For the previous model, $freq_{day}$ was taken into account too, but this lead to over-performance of the prospected market share.

Even though little data was available, the market share model still succeeds to calculate a satisfying average prospected market share for the market share model without $freq_{day}$ for the month September 2015. However, there is a doubt if the market share model for this market can be validated if more data would be available from other months because there is little variation for the value for $freq_{week}$. The market share model with all airline related variables could not be validated since the difference between the actual and prospected market share is too high.

7.4. Concluding remarks about testing the model

The various market share models were tested in this section. First, a sensitivity analysis is carried out. Therefore, one of the significant variables are set to its minimum and maximum value and the remaining variables are kept at its mean. Then the change in market share is calculated. The independent variable $freq_{day}$ was found to be the variable that triggers the highest changes in market share. $Price_{day}$ and $freq_{week}$ also achieved changes in market share but these changes were only once higher than 4%.

The verification was already accomplished in the results section since there, the sign of the β-coefficients of each corresponding independent variable is analyzed and discussed.

Also, a validation was performed for the market share models with all airline related independent variables and without $freq_{day}$ after liberalization. No validation data was available for the model before liberalization. Also only validation data for September 2015 was available which means that the market share model with monthly dummy variables could not be tested.

In the end, the market share models for the combined market and LIS-PDL-LIS could be validated for the market share model with all airline related variables. The models of the market OPO-PDL-OPO for the market share model with all airline related variables and of the combined market for the market share model without $freq_{day}$ did not perform well in the validation. The market share models should not be preferred over the market share models that perform very well under the validation circumstances but still perform satisfactorily.

8

Conclusions, Reflection, and Recommendations

Conclusions are drawn from the results of the market share model. These conclusion can be found in Section 8.1. Then, a reflection is made on the limitation of the market share model in Section 8.2. Furthermore, recommendations are given to further improve and expand the market share model which can be found in Section 8.3.

8.1. Conclusion

In this section conclusions are drawn from the results of the developed market share model. This is done using the previously defined research questions from Section 3.1 that build the structure of the project. The market share is applied to the Azorean archipelago because it is a good example of a newly liberalized low-demand market.

How can the market share of airlines in a newly liberalized low-demand market be modeled?

The methodology of the generalized linear model with logit link function, binomial family, and robust standard errors was chosen to be appropriated for the market share model. These calculations are implemented in the statistic software STATA®, which was used during this project. The software was able to process the input data that was provided to the model.

The input data included airline characteristics of the airlines in the market for each day as the independent variables and market share as dependent variable. The β-coefficients for each corresponding airline characteristic were calculated. These coefficients could then be used for analyzing the impact of each independent vairable and to compare the market share models before and after liberalization.

Which attributes can be identified that have a significant impact on the market share in the described market?

The selection of the attributes is based on literature study. The airline characteristics under investigation are: ticket price the day prior departure and two month before departure, frequency per day and per week, time-points (measure of flights within a preferred time window) and on-time performance.

Additionally, market share models were developed by adding monthly dummy variables to the set of independent variables.

However, not all variables are identified to be significant in each market. Since there is a different market share model for each market, different attributes are found to be significant. All significant airline attributes for each market can be found in Table 8.1.

Table 8.1: Significant independent variables

	after liberalization			before liberalization
	LIS-OPO	LIS-PDL-LIS	OPO-PDL-OPO	LIS-PDL-LIS
all	$freq_{day}$ constant	$price_{day}$ $freq_{day}$ constant	$freq_{day}$ $freq_{week}$	$freq_{day}$ constant
without $freq_{day}$	$price_{2m}$ time-points constant	$price_{day}$ time-points constant	$freq_{week}$ constant	$price_{2m}$ $freq_{week}$ constant
monthly dummies	$freq_{day}$ constant	$freq_{day}$ May June July constant	$price_{day}$ $freq_{day}$ June July constant	

How does the influence of the attributes change due to the liberalization?

The case between Ponta Delgada and Lisbon was used to point out the change in impact on market share due to liberalization. Frequency per day is significant in both situations for the market share model with all airline realted variables, as can be seen in Table 8.1. Frequency per day gained importance whereas the impact of the constant decreased due to liberalization. In the impact analysis, this impact on market share can be expressed in numbers. Frequency per day took over 37.88% of the contribution to market share before liberalization and 47.15% after liberalization. For the constant, this changes from 62.12% to 43.1% due to liberalization.

For the situation after liberalization, the independent variable $price_{day}$ was also significant and stayed the same when $freq_{day}$ was removed from the model, but the impact increased from 9.75% to 36% due to deleting that independent varibale. Comparing the remaining significant variables for the model without $freq_{day}$, before liberalization, $price_{2m}$ and $freq_{week}$ were significant whereas after liberalization $price_{day}$ and $time\_points$ appear to be significant.

What is the probability that passengers choose for a specific airline in the described market in the future?

For each market, an equation can be developed that can be used to calculate the market share in the future. It consists of the calculated β-coefficients of that market. The airline characteristics of the future situation serve as independent variables for the equation. The future market share of each market can be calculated with these input data. From the validation it could be concluded that the equations for the combined markets and the market between Ponta Delgada and Lisbon perform satisfactory.

Main research question: ***What is the impact of liberalization on airline market share in a newly liberalized low-demand market?***

By answering all the sub-research questions the main research question can be answered too. The independent variable $freq_{day}$ was significant in all market share models (if included) and triggered the biggest changes in market share. If an airline would like to be competitive in the market, the improvement of $freq_{day}$ should be the focus of the scheduling strategy. Due to liberalization, the importance of $freq_{day}$ increases and the $price_{day}$ became significant, while it was not significant before for the market share model with all airline spcific variables. For the market share model without $freq_{day}$, $price_{2m}$ and $freq_{week}$ were significant before liberalization but after liberalization $price_{day}$ and $time\_points$ were found to be significant.

Additional findings

Market share models were developed with monthly dummy variables for the situation after liberalization. It was found that for the route LIS-PDL-LIS a positive timely effect could be found towards the summer whereas in the route OPO-PDL-OPO the effect was negative.

Additionally, the markets between Ponta Delgada and Porto as well as the combined market were analyzed. The market share model to Porto suffered from limitations in the calibration and validation data. Therefore, the results that state that both $freq_{day}$, as well as $freq_{week}$ (but no constant) were significant the case with all airline-related independent variables. When $freq_{day}$ was removed, the constant appeared to take over a major part of the impact on market share.

For the combined market share model, the monthly dummies showed no effect on the market share model. Furthermore, when all airline-related variables are used, $freq_{day}$ and the constant share both approximately half of the impact, However, when $freq_{day}$ is removed the impact is approximately evenly divided over $price_{2m}$, $time\_points$, and the constant.

8.2. Limitations and Reflection

First of all, there are several weaknesses of the model that need to be pointed out. One issue is that the input data has to be taken with caution. The data of the independent variable on_time of SATA and TAP was calculated with the use of the passenger data. For this, the number of flights that were delayed for more than 15 minutes were divided by the total number of flights. Then, this share of flights that were too late is transformed in the on-time performance. One problem is that on-time performance of the LCC's is not available for the Azorean market. The airline only publishes on-time performance data for their entire network, which means that the actual on-time performance that has impact on the market share in this market could be different.

Additionally, the market share with which the market share model is calibrated is an approximation. The daily passenger numbers of the LCC's are not known - only the monthly numbers are. This means that the total daily passenger numbers are approximated from a graph found in the airport report from Ponta Delgada airport for each month. Then, the passenger numbers, that do not belong to the route under consideration, are subtracted from the total daily passenger number that are handled on the airport. Because Ryanair flies on both routes under consideration, the passenger numbers of the route that does not belong to the market are spread over the month by taking weekly seasonality into account. These passenger values are not equal to the actual passenger values that flew on that route that day, but because this data is not available, these approximations have to be made.

Furthermore, it is a major problem of the market share model that the calibration data has little variation. Therefore, little changes in the independent variables trigger big changes in the results. With a larger data set that runs all year around, higher variations are expected and a model with monthly dummy variables can be created that can be validated.

Another problem in the model is correlation. In the first model, all airline-related variables are considered even though $freq_{day}$ is highly correlated to market share. Therefore, the second market share model is created where $freq_{day}$ is deleted. For all other variables, a variable is deleted from the model if the correlation coefficient is above 0.7 Hinkle et al. [19]. However, even if the correlation coefficient is lower than 0.7 it does not mean that these variables do not have effect on each other. It is assumed that the effect is only significant above 0.7 but one should recognize the effect that is remaining.

Also, the network effect should not be disregarded. A part of the passengers will not only travel from Ponta Delgada to the Portuguese mainland, but will continue to another destination. When SATA is the carrier of the connecting flight, the passenger can reach the remaining 8 Azorean islands or an airport in North America. When TAP is the connecting carrier, the passenger could travel to the whole variety of destinations within the STAR alliance.

One thing is particularly interesting: SATA is forced to provide connecting flights to the remaining Azorean islands for free even if the passengers use an LCC to arrive to Ponta Delgada or any other gateway of the Azores. This means that even the costs of taxes are billed to SATA and not to the passenger or the LCC. This regulation rests heavily on the airline that is already in a bad financial situation.

The Azores is taken as a case study for a low-demand newly liberalized market. However, it is unsure if the market share model can be applied to other markets too. The selection of the independent variables has the application on the Azores in mind which resulted in the fact that for example flight time was not taken into account. The assumption that the case study is representative for other low-demand newly liberalized market is made, but not tested yet.

8.3. Recommendation

In the previous section it could be seen that the developed market share model can answer the research questions that are selected from this project. However, the research also has some limitations that were addressed above that could be resolved. Improvements can be made so that the model is able to produce better results. In this section, ideas are given how the model can be expanded or improved so that the limitations of the market share model are not an issue any more.

Improve the input data

The model mainly suffers due to the quality of the data that is provided. The market share data is based on approximations and the ticket price data is taken from accumulated ticket sales per day which also included employee tickets, group tickets, or promotion tickets that are reserved for inhabitants of the Azores islands only. Also, the calibration data the model is based on four months. If this could be extended, the market share model would deliver better results. Especially for the market between Ponta Delgada and Porto, this would be the key to create a model with β-coefficients that have the appropriate sign.

Data collection is a time-consuming task when there is only little information available for the market under consideration. The market share model of the route between Ponta Delgada and Lisbon before liberalization has shown that with daily passenger flow data, ticket price data, and on-time performance data of all the airlines in the market, the market share model outputs results that are very satisfying.

Investigate the effect on stop-over and travel time on the full Azorean market

The market share model is made for competitors on the exact same route so travel time and stop-over possibilities are not taken into account. It could be a possible expansion for the model to include these two attributes. Then the passengers could be taken into account that fly to Lisbon or Porto with a stop-over.

Also if stop-overs are taken into account, the market between all the Azorean island and the mainland can be investigated. The archipelago consists of nine different islands that are all inhabited. It could be valuable to better understand the passenger booking behavior of the region in this way.

Apply the market share model to another recently liberalized low-demand market

In the scope of this thesis it was not possible to test if the market share model could also be applied to other newly liberalized low-demand markets or even markets that have high demand. Especially for countries that have many islands, like Norway or Greece, the market share model can be applied to. The market share model could also be applied to Madeira, another Portuguese island that is already liberalized for a longer period of time. Then, conclusions can be drawn about if the passenger behavior is similar to the one captured in the Azorean market.

The author is convinced that the methodology can be applied to any market with any kind of attribute that are not collinear to each other and can be expressed as a numeric value. Therefore, any kind of situation can be analyzed.

Liberalization of a market is just one example. Also the impact of war, closure or opening of a new airport, and the difference between leisure and business passenger on passenger booking behavior can be investigated.

A

Correlation matricies

A.1. Correlation matricies of the markets in the situation after liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time	may	june	july
market_sh~01	1.0000									
price_day	-0.2789	1.0000								
price_2m	-0.3319	0.3461	1.0000							
freq_day	0.8687	-0.1913	-0.4197	1.0000						
freq_week	0.1875	-0.1281	-0.3125	0.2061	1.0000					
time_points	0.3151	-0.1775	-0.1160	0.2972	0.1562	1.0000				
on_time	0.1126	-0.2145	-0.1250	0.0789	-0.1031	-0.1839	1.0000			
may	0.1105	-0.1931	-0.0951	0.0824	-0.0967	-0.1900	0.9966	1.0000		
june	0.1611	0.0798	0.0487	0.1765	0.3279	0.0586	-0.3341	-0.2740	1.0000	
july	0.1625	-0.2965	-0.4674	0.0814	0.1551	0.1516	-0.2198	-0.2889	-0.2189	1.0000

Figure A.1: Correlation matrix of the combined market after liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time	may	june	july
market_sh~01	1.0000									
price_day	-0.3559	1.0000								
price_2m	-0.0116	-0.1332	1.0000							
freq_day	0.8194	-0.3152	0.0056	1.0000						
freq_week	0.1736	-0.2885	-0.1091	0.2333	1.0000					
time_points	0.4801	-0.3079	-0.0235	0.5804	0.0983	1.0000				
on_time	-0.2757	0.4245	0.0805	-0.0384	-0.2035	-0.2083	1.0000			
may	-0.0668	-0.1829	0.5475	-0.0533	-0.0703	-0.1619	-0.1559	1.0000		
june	0.1199	-0.1066	-0.1481	0.0712	0.2836	0.2132	-0.1598	-0.3691	1.0000	
july	0.2087	-0.1808	-0.3422	0.0214	0.0342	0.1492	-0.6757	-0.2871	-0.2943	1.0000

Figure A.2: Correlation matrix of the route LIS-PDL-LIS after liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time	may	june	july
market_sh~01	1.0000									
price_day	0.0323	1.0000								
price_2m	-0.0824	-0.2290	1.0000							
freq_day	0.8194	0.0257	-0.0629	1.0000						
freq_week	0.0480	-0.1789	-0.0917	0.1663	1.0000					
time_points	-0.0481	0.0776	0.0414	-0.0422	0.0695	1.0000				
on_time	-0.1345	0.0977	0.4522	0.0412	-0.0412	0.0092	1.0000			
may	-0.0771	-0.2882	0.5526	0.0180	-0.0572	-0.0149	0.7106	1.0000		
june	-0.0307	-0.1163	-0.1833	-0.0385	0.1264	-0.0030	-0.3812	-0.3827	1.0000	
july	0.1579	-0.1390	-0.2474	-0.0192	-0.0509	-0.0142	-0.7326	-0.2834	-0.3129	1.0000

Figure A.3: Correlation matrix of the route LIS-PDL after liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time	may	june	july
market_sh~01	1.0000									
price_day	-0.1336	1.0000								
price_2m	0.1312	-0.1558	1.0000							
freq_day	0.8316	-0.0768	0.2250	1.0000						
freq_week	0.1057	-0.1244	-0.0283	0.1306	1.0000					
time_points	0.4150	-0.2157	0.0778	0.4803	0.0113	1.0000				
on_time	0.1223	0.1142	-0.4827	-0.0679	0.1490	0.0446	1.0000			
may	-0.0221	-0.2001	0.3762	0.0442	-0.0459	-0.0471	-0.7625	1.0000		
june	-0.0649	-0.0196	-0.0135	-0.0624	0.0895	0.0351	0.0068	-0.3447	1.0000	
july	0.1613	0.0242	-0.4046	-0.0543	0.1603	0.0244	0.8526	-0.3208	-0.3626	1.0000

Figure A.4: Correlation matrix of the route PDL-LIS after liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time	may	june	july
market_sh~01	1.0000									
price_day	-0.0179	1.0000								
price_2m	0.0825	0.2960	1.0000							
freq_day	0.5175	0.1323	-0.1438	1.0000						
freq_week	-0.3758	-0.0441	-0.2455	-0.1473	1.0000					
time_points	0.0392	0.0336	-0.1664	-0.1253	0.1979	1.0000				
on_time	0.1895	-0.0785	-0.0900	0.0673	-0.1407	-0.0755	1.0000			
may	0.4074	0.0774	0.0768	0.1179	-0.3382	-0.0752	0.9186	1.0000		
june	-0.2601	0.4022	0.3987	0.0144	0.1725	-0.1287	-0.3487	-0.2854	1.0000	
july	-0.5076	-0.4556	-0.4777	-0.1430	0.4680	0.0515	0.0205	-0.3651	-0.3127	1.0000

Figure A.5: Correlation matrix of the route OPO-PDL-OPO after liberalization

A.2. Correlation matricies of the markets in the situation before liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time
market_sh~01	1.0000						
price_day	-0.1593	1.0000					
price_2m	-0.1253	-0.1657	1.0000				
freq_day	0.5450	-0.2447	0.0224	1.0000			
freq_week	0.3333	-0.4249	0.2301	0.5731	1.0000		
time_points	-0.0423	-0.0514	0.0135	-0.0441	-0.0165	1.0000	
on_time	0.1825	-0.1282	0.0413	0.0634	-0.0786	0.1042	1.0000

Figure A.6: Correlation matrix of the route LIS-PDL-LIS before liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time
market_sh~01	1.0000						
price_day	0.1163	1.0000					
price_2m	-0.0640	-0.1996	1.0000				
freq_day	0.6490	-0.1215	0.0062	1.0000			
freq_week	0.3863	-0.0012	-0.0893	0.3273	1.0000		
time_points	0.3063	-0.0743	-0.1755	0.2804	0.2261	1.0000	
on_time	0.1041	0.0117	0.1222	0.0897	-0.1396	0.0432	1.0000

Figure A.7: Correlation matrix of the route LIS-PDL before liberalization

	marke~01	price_~y	price_2m	freq_day	freq_w~k	time_p~s	on_time
market_sh~01	1.0000						
price_day	-0.4001	1.0000					
price_2m	0.1123	-0.1004	1.0000				
freq_day	0.7878	-0.4107	0.1962	1.0000			
freq_week	0.5657	-0.6046	0.2135	0.6683	1.0000		
time_points	0.3325	-0.1274	0.1538	0.3807	0.2249	1.0000	
on_time	0.0497	0.1802	0.4544	0.0291	-0.0321	0.0093	1.0000

Figure A.8: Correlation matrix of the route PDL-LIS before liberalization

B

Results of the market share model for the situation after liberalization

B.1. Results of the market share model with all airline-related variables

```
Fractional logistic regression                  Number of obs     =         64
                                                Wald chi2(1)      =     267.61
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -41.817863               Pseudo R2         =     0.0198
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	4.429658	.2707803	16.36	0.000	3.898939	4.960378
_cons	-2.094553	.0981511	-21.34	0.000	-2.286925	-1.90218

Figure B.1: Results of the combined market for the the market share model with all airline-related variables after liberalization

```
Fractional logistic regression                  Number of obs     =        102
                                                Wald chi2(2)      =     305.40
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -66.106841               Pseudo R2         =     0.0308

------------------------------------------------------------------------------
             |               Robust
market_share01 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   price_day |  -.1299645   .0656538    -1.98   0.048    -.2586436   -.0012854
    freq_day |   4.666971   .3195544    14.60   0.000     4.040655    5.293286
       _cons |  -1.896145   .1661809   -11.41   0.000    -2.221854   -1.570437
------------------------------------------------------------------------------
```

Figure B.2: Results of the route LIS-PDL-LIS for the the market share model with all airline-related variables after liberalization

```
Fractional logistic regression                  Number of obs     =        101
                                                Wald chi2(1)      =     194.79
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -65.845324               Pseudo R2         =     0.0326

------------------------------------------------------------------------------
             |               Robust
market_share01 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    freq_day |   5.063026   .3627641    13.96   0.000     4.352022    5.774031
       _cons |  -2.159993   .1201189   -17.98   0.000    -2.395421   -1.924564
------------------------------------------------------------------------------
```

Figure B.3: Results of the route LIS-PDL for the the market share model with all airline-related variables after liberalization

```
Fractional logistic regression                  Number of obs     =        107
                                                Wald chi2(1)      =     201.36
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -70.700894               Pseudo R2         =     0.0283

------------------------------------------------------------------------------
             |               Robust
market_share01 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    freq_day |    4.49715   .3169182    14.19   0.000     3.876002    5.118299
       _cons |  -2.006463   .1163648   -17.24   0.000    -2.234534   -1.778392
------------------------------------------------------------------------------
```

Figure B.4: Results of the route PDL-LIS for the the market share model with all airline-related variables after liberalization

```
Fractional logistic regression                  Number of obs     =         56
                                                Wald chi2(2)      =      25.15
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -38.065056               Pseudo R2         =     0.0046
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	3.193491	.9812249	3.25	0.001	1.270326	5.116657
freq_week	-2.474369	.8064961	-3.07	0.002	-4.055072	-.8936658
_cons	-.6133631	.6974902	-0.88	0.379	-1.980419	.7536926

Figure B.5: Results of the route OPO-PDL-OPO for the the market share model with all airline-related variables after liberalization

B.2. Results of the market share model without $freq_{day}$ as independent variable

```
Fractional logistic regression                  Number of obs     =         64
                                                Wald chi2(2)      =      14.18
                                                Prob > chi2       =     0.0008
Log pseudolikelihood = -42.456587               Pseudo R2         =     0.0048
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
price_2m	-.1283861	.0562066	-2.28	0.022	-.238549	-.0182231
time_points	.6528942	.2873706	2.27	0.023	.0896582	1.21613
_cons	-.4903496	.1729142	-2.84	0.005	-.8292551	-.1514441

Figure B.6: Results of the combined market for the the market share model without $freq_{day}$ after liberalization

```
Fractional logistic regression                  Number of obs     =        102
                                                Wald chi2(2)      =      62.05
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -67.373647               Pseudo R2         =     0.0122
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
price_day	-.2470185	.0824073	-3.00	0.003	-.4085339	-.0855031
time_points	.797951	.1572639	5.07	0.000	.4897194	1.106183
_cons	-.6075014	.1788281	-3.40	0.001	-.9579981	-.2570047

Figure B.7: Results of the route LIS-PDL-LIS for the the market share model without $freq_{day}$ after liberalization

```
Fractional logistic regression                  Number of obs     =        107
                                                Wald chi2(1)      =      23.05
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -72.256182               Pseudo R2         =     0.0069
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
time_points	.5725947	.1192629	4.80	0.000	.3388438	.8063456
_cons	-.5792125	.069948	-8.28	0.000	-.7163081	-.4421169

Figure B.8: Results of the route PDL-LIS for the the market share model without $freq_{day}$ after liberalization

```
Fractional logistic regression                  Number of obs     =         56
                                                Wald chi2(1)      =      12.22
                                                Prob > chi2       =     0.0005
Log pseudolikelihood = -38.173567               Pseudo R2         =     0.0018
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_week	-3.042547	.8704811	-3.50	0.000	-4.748659	-1.336435
_cons	1.272385	.4389351	2.90	0.004	.4120884	2.132682

Figure B.9: Results of the route OPO-PDL-OPO for the the market share model without $freq_{day}$ after liberalization

B.3. Results of the market share model adding monthly dummies

```
Fractional logistic regression                  Number of obs     =        102
                                                Wald chi2(4)      =     436.23
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -65.929281               Pseudo R2         =     0.0334
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	4.877456	.2560996	19.05	0.000	4.37551	5.379402
may	.166651	.0727391	2.29	0.022	.0240849	.3092171
june	.2488923	.0754783	3.30	0.001	.1009576	.396827
july	.3957976	.0939927	4.21	0.000	.2115753	.5800199
_cons	-2.335798	.099323	-23.52	0.000	-2.530468	-2.141129

Figure B.10: Results of the route LIS-PDL-LIS for the the market share model with monthly dummies after liberalization

```
Fractional logistic regression                  Number of obs     =         56
                                                Wald chi2(3)      =      98.39
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -37.927459               Pseudo R2         =     0.0082
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	2.964868	.5958212	4.98	0.000	1.79708	4.132656
june	-.3006099	.0510272	-5.89	0.000	-.4006214	-.2005984
july	-.3480477	.049025	-7.10	0.000	-.4441349	-.2519605
_cons	-1.611542	.3012522	-5.35	0.000	-2.201986	-1.021099

Figure B.11: Results of the route OPO-PDL-OPO for the the market share model with monthly dummies after liberalization

C

Results of the market share model for the situation before liberalization

C.1. Results of the market share model with all airline-related variables

```
Fractional logistic regression                  Number of obs     =         71
                                                Wald chi2(1)      =      14.35
                                                Prob > chi2       =     0.0002
Log pseudolikelihood = -38.321051               Pseudo R2         =     0.0053
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	3.492359	.9219395	3.79	0.000	1.685391	5.299327
_cons	-1.336337	.6759533	-1.98	0.048	-2.661181	-.0114928

Figure C.1: Results of the route LIS-PDL-LIS for the the market share model with all airline-related variables before liberalization

```
Fractional logistic regression                  Number of obs     =         68
                                                Wald chi2(3)      =      65.43
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -36.247852               Pseudo R2         =     0.0065
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
price_day	.1524253	.0735406	2.07	0.038	.0082883	.2965623
freq_day	3.085768	.4242054	7.27	0.000	2.254341	3.917195
freq_week	1.924568	.7290404	2.64	0.008	.4956746	3.35346
_cons	-2.811398	.6876869	-4.09	0.000	-4.159239	-1.463556

Figure C.2: Results of the route LIS-PDL for the the market share model with all airline-related variables before liberalization

```
Fractional logistic regression                  Number of obs     =         64
                                                Wald chi2(1)      =     129.42
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -37.391801               Pseudo R2         =     0.0194
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_day	4.15315	.3650643	11.38	0.000	3.437637	4.868663
_cons	-1.955001	.2609662	-7.49	0.000	-2.466485	-1.443517

Figure C.3: Results of the route PDL-LIS for the the market share model with all airline-related variables before liberalization

C.2. Results of the market share model without $freq_{day}$ as independent variable

```
Fractional logistic regression                  Number of obs     =         71
                                                Wald chi2(2)      =      14.64
                                                Prob > chi2       =     0.0007
Log pseudolikelihood = -38.419737               Pseudo R2         =     0.0027
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
price_2m	-.0546871	.0241964	-2.26	0.024	-.1021112	-.007263
freq_week	5.426765	1.755123	3.09	0.002	1.986787	8.866742
_cons	-3.203385	1.46189	-2.19	0.028	-6.068636	-.3381331

Figure C.4: Results of the route LIS-PDL-LIS for the the market share model without $freq_{day}$ before liberalization

```
Fractional logistic regression                  Number of obs     =         68
                                                Wald chi2(2)      =      15.40
                                                Prob > chi2       =     0.0005
Log pseudolikelihood = -36.408567               Pseudo R2         =     0.0021
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
price_day	.0954306	.0894935	1.07	0.286	-.0799734	.2708345
freq_week	3.877888	1.053066	3.68	0.000	1.813917	5.941859
_cons	-2.093533	.8612427	-2.43	0.015	-3.781538	-.405528

Figure C.5: Results of the route LIS-PDL for the the market share model without $freq_{day}$ before liberalization

```
Fractional logistic regression                  Number of obs     =         64
                                                Wald chi2(2)      =      42.17
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -37.677862               Pseudo R2         =     0.0119
```

market_share01	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
freq_week	11.25524	1.791263	6.28	0.000	7.744425	14.76605
time_points	.4211409	.1503265	2.80	0.005	.1265063	.7157755
_cons	-8.738457	1.532479	-5.70	0.000	-11.74206	-5.734853

Figure C.6: Results of the route PDL-LIS for the the market share model without $freq_{day}$ before liberalization

Bibliography

[1] Adler, N. Competition in a deregulated air transportation market. *European Journal of Operational Research*, 129(2):337–345, 2001.

[2] Adler, N., Pels, E., and Nash, C. High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. *Transportation Research Part B: Methodological*, 44(7):812–833, 2010.

[3] Adler, N., Fu, X., Oum, T. H., and Yu, C. Air transport liberalization and airport slot allocation: The case of the Northeast Asian transport market. *Transportation Research Part A: Policy and Practice*, 62:3–19, 2014.

[4] Barnhart, C. and Smith, B. *Quantitative Problem Solving Methods in the Airline Industry*. Springer, 2012.

[5] Barone, J. The Azores: An untamed island paradise only a few hours by air from North America, 2015. URL `https://www.washingtonpost.com/lifestyle/travel/`.

[6] Belobaba, P., Odoni, A., and Barnhart, C. *The Global Airline Industry*. Wiley, 2009.

[7] Ben-Akiva, M. and Lerman, S. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Massachusetts, 1985.

[8] Berkson, J. Application of the logistic function to bio-assay. *Journal of the American Statistical Association*, 39(227):357–365, 1944.

[9] Bruin, J. Stata Annotated Output Correlation, 2011. URL `http://www.ats.ucla.edu/stat/stata/output/stata{_}corr{_}output.htm`.

[10] Carrier, E. *Modeling the Choice of an Airline Itinerary and Fare Product Using Booking and Seat Availability Data*. PhD thesis, Massachusetts Institute of Technology, 2008. URL `http://agifors.org/award/submissions2008/Carrier{_}paper.pdf`.

[11] Cohas, F. J., Belobaba, P. P., and Simpson, R. W. Competitive fare and frequency effects in airport market share modeling. *Journal of Air Transport Management*, 2(I):33–45, 1995.

[12] Dobson, A. J. and Barnett, A. *An Introduction to Generalized Linear Models*. 2008.

[13] Domencich, T. and McFadden, D. *Urban travel demand - a behavioral analysis*. North-Holland Publishing Company, 1975.

[14] Dresner, M., Lin, J. S. C., and Windle, R. The impact of low-cost carriers on airport and route competition, 1996.

[15] European Court of Justice. Case 283/06, 2006.

[16] Flightglobal. Ryanair to open Azores base, 2014. URL `https://www.flightglobal.com/news/articles/ryanair-to-open-azores-base-406794/`.

[17] Fu, X., Oum, T., and Zhang, A. Air transport liberalization and its impacts on airline competition and air passenger traffic. *Transportation journal*, 49:24, 2010.

[18] Garrow, L. *Discrete Choice Modelling and Air Travel Demand*. Ashgate, 2010.

[19] Hinkle, D., Wiersma, W., and Jurs, S. *Applied Statistics for the Behavioral Sciences*. Wadsworth Publishing, 5 edition, 2003.

[20] Hsu, C. I. and Wen, Y. H. Airline flight frequency determination in response to competitive interactions using fuzzy logic. *Mathematical and Computer Modelling*, 42:1207–1224, 2005.

[21] Hsu, C. W., Lee, Y., and Liao, C. H. Competition between high-speed and conventional rail systems: A game theoretical approach. *Expert Systems with Applications*, 37(4):3162–3170, 2010.

[22] Janić, M. Liberalisation of European aviation: analysis and modelling of the airline behaviour. *Journal of Air Transport Management*, 3(4):167–180, 1997.

[23] Johnson, P. E. GLM #2 (version 2) Residuals and analysis of fit. Technical report, 2016.

[24] Levy, S. and Panou, C. To Travel or Not To Travel: a Study of Islanders'Trips To the Mainland. *12th WCTR, July 11-15, 2010 – Lisbon, Portugal*, pages 1–15, 2010.

[25] McFadden, D. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pages 105–142. 1973.

[26] McFadden, D., Tye, W., and Train, K. *An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model*. Institute of Transportation Studies, University of California, 1977.

[27] Nash, J. The Bargaining Problem. *Journal of the Econometric Society*, 18(2):155–162, 1950.

[28] Nash, J. Non-Cooperative Games. *Annals of Mathematics*, 54:286–295, 1951.

[29] Nelder, A. J. a. and Wedderburn, R. W. M. Generalized Linear Models. *J. R. Statist. Soc. A.*, 135 (3):370–384, 1972.

[30] O'Connell, J. F. and Williams, G. Passengers' perceptions of low cost airlines and full service carriers: A case study involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines. *Journal of Air Transport Management*, 11:259–272, 2005.

[31] Olsson, U. *Generalized linear model - an applied approach*. Studentlitteratur, Lund, Sweden, 2002.

[32] Papke, L. E. and Wooldridge, J. M. Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632, 1996.

[33] Pearson, K. Contributions to the mathematical theory of evolution. *Philosophical Transactions A*, 185(A):71–110, 1894.

[34] Pearson, K. On the criterion that a given system of deviantions from the probable in the case of a correlated system of variables in such that i can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302):157–175, 1900.

[35] Pels, E. Airline network competition: Full-service airlines, low-cost airlines and long-haul markets. *Research in Transportation Economics*, 24(1):68–74, 2009.

[36] Prince, J. and Simon, D. Multimarket contact and service quality: Evidence from on-time performance in the U.S. airline industry. *Academy of Management Journal*, 52(2):336–354, 2009.

[37] Proussaloglou, K. and Koppelman, F. Air carrier demand an analysis of market share determinate.pdf. *Transportation*, pages 371–388, 1995.

[38] Raposo, S. SATA, 2014.

[39] Román, C., Espino, R., Martín, J. C., Betancor, O., and Nombela, G. Analyzing mobility in peripheral regions of the European Union: The case of Canarias-Madeira-Azores. *Networks and Spatial Economics*, 8:141–160, 2008.

[40] StataCorp. 2013. Stata 13 Base Reference Manual - robust, . URL http://www.stata.com/manuals13/p{_}robust.pdf.

[41] StataCorp. 2013. Stata 13 Base Reference Manual - correlate, . URL http://www.stata.com/manuals13/rcorrelate.pdf.

[42] StataCorp. 2014. Stata 14 Base Reference Manual - fracreg. URL http://www.stata.com/manuals14/rfracreg.pdf.

[43] Suzuki, Y. The relationship between on-time performance and airline market share: A new approach. *Transportation Research Part E: Logistics and Transportation Review*, 36:139–154, 2000.

[44] Suzuki, Y., Tyworth, J. E., and Novack, R. a. Airline market share and customer service quality: a reference-dependent model. *Transportation Research Part A: Policy and Practice*, 35:773–788, 2001.

[45] UCLA: Statistical Consulting Group. FAQ: What are pseudo R-squareds? URL http://www.ats.ucla.edu/stat/mult{_}pkg/faq/general/Psuedo{_}RSquareds.htm.

[46] UCLA: Statistical Consulting Group. FAQ: What are pseudo R-squareds? URL http://www.ats.ucla.edu/stat/mult{_}pkg/faq/general/Psuedo{_}RSquareds.htm.

[47] Von Neumann, J. *Zur Theorie der Gesellschaftsspiele*, volume 100. Springer, 1928.

[48] Wasserman, L. All of Statistics : A Concise Course in Statistical Inference Brief Contents. *Simulation*, C:461, 2004.

[49] Wei, W. and Hansen, M. Cost economics of aircraft size. *Journal of Transport Economics and Policy*, 37(2):279–296, 2003.

[50] Wei, W. and Hansen, M. Impact of aircraft size and seat availability on airlines' demand and market share in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review*, 41:315–327, 2005.

[51] Wei, W. and Hansen, M. Airlines' competition in aircraft size and service frequency in duopoly markets. *Transportation Research Part E: Logistics and Transportation Review*, 43:409–424, 2007.

[52] Zhang, H., Su, Y., Peng, L., and Yao, D. A review of game theory applications in transportation analysis. *Proceedings of ICCIA 2010 - 2010 International Conference on Computer and Information Application*, pages 152–157, 2010.