



UiT The Arctic University of Norway

School of Business and Economics

Pricing strategies of airline companies

Discovering the reasons behind ticket price variation on Norwegian airline market

Sofia Grabovskaia

Master's thesis in Economics, SOK-3901, November 2020

Acknowledgements

First and foremost, I would like to express gratitude to my supervisor Associate Professor Jinghua Xie of the School of Business and Economics at University of Tromsø for support of my research question, valuable comments, suggestions and feedback throughout the working process.

My appreciation also goes to the staff at the School of Business and Economics. Lectures at the master's program have been very interesting and engaging.

Finally, a thank you goes to my family and friends for the incredible support throughout my years of study.

Tromsø, November 2020

Sofiia Grabovskaia

Abstract

Price of good or service is to some extent a reflection of company's underlying strategies. On the airline market, price is determined by various factors such as type of carrier, season, flight and fare characteristics, as well as goals and visions of the specific airline company.

This thesis contains an empirical analysis of pricing strategies of airline companies on domestic market in Norway. The research uses primary data on airline fare prices collected directly from websites of two established airlines on Norwegian market, network carrier Scandinavian Airlines (SAS) and low-cost carrier Norwegian Air Shuttle (Norwegian) during the period 8th August 2020 – 10th September 2020. Data sample consists of all flights operated by SAS and Norwegian between five most trafficked airports in Norway according to SSB statistics, with departures on Monday 7th and Friday 11th September 2020. The airports included in the sample are Oslo Gardermoen, Bergen Flesland, Trondheim Værnes, Stavanger Sola and Tromsø Langnes. The analysis uses multiple linear regression with the aim to determine if such characteristics as day of the week when flight departs, flight duration, time to flight departure, market share of the airline on route, competition on departure and time of departure impact the ticket prices on average according to the collected data.

The research is carried out during the times of COVID-19 pandemic which leads to more complicated situation on the airline market. It was, therefore, expected that some natural assumptions on strategies of airlines may not be supported by the data.

The key results are that our data supports that, average fare prices raise with flight length, as well as that average prices increase over time prior to departure day. Moreover, the data provides the evidence that market share on route, competition on departure and time of departure explain some variation in average ticket price in both positive and negative direction depending on the airline company. Finally, it is supported by the data that there are, to some extent, different patterns for pricing of flights which depart on Monday 7th September 2020 and Friday 11th September 2020.

Key words: Airlines, Pricing, SAS, Norwegian

Abbreviations

AB/AS	Stock-based company (Norwegian: Aksjeselskap)
COVID-19	COronaVirus Disease 2019
GLS	Generalized Least Squares
IT	Information Technology
NA	Not available
NOK	Norwegian krones
OLS	Ordinary Least Squares
SAS	Scandinavian Airlines System (Scandinavian Airlines)
SSB	Statistics Norway (Norwegian: Statistisk sentralbyrå)
UK	United Kingdom
US	United States
VIF	Variance Inflation Factor

Table of Contents

1. Introduction.....	1
1.1 Revenue management	3
1.2 Price discrimination.....	4
1.3 Price discrimination and the airline market.....	5
1.4 Previous research.....	5
2. The airline market.....	7
2.1 Network airlines	7
2.2 Low cost airlines	8
2.3 Airline market in Norway	9
2.4 Scandinavian Airlines	10
2.5 Norwegian Air Shuttle	11
3. Method and data.....	13
3.1 Linear regression analysis	13
3.2 Pooled OLS, Fixed Effects and Random Effects models	17
3.3 GLS estimation.....	18
3.4 Data collection process.....	19
4. Method implementation and results.....	25
4.1 Choice of variables.....	25
4.2 The general analysis	29
4.3 The “Economy subset” analysis.....	35
4.4 “Norwegian” and “SAS” analysis.....	37
5. Concluding discussion	43
6. List of references.....	47
7. Appendix.....	53

List of tables

Table 1: Choice of the model based on results of statistical tests.....	18
Table 2: Flight movements in Norway – departure and arrivals.....	20
Table 3: Airports in Norway with the highest traffic, 4 th quarter 2019	21
Table 4: Traffic between the chosen Norwegian airports, 4 th quarter 2019	21
Table 5: Overview of ticket types and service available for domestic travels.....	23
Table 6: Overview of ticket types and service available for domestic travels.....	23
Table 7: Suggested variables for regression model.....	28
Table 8: Variables overview.....	30
Table 9: Results of Chow and Hausman tests for the general models.....	31
Table 10: Results of Breusch-Pagan and Durbin-Watson tests for the general models.....	32
Table 11: VIF values for the general models.....	32
Table 12: Summary of the general regression results.....	34
Table 13: Summary of the “Economy subset” regression results.....	37
Table 14: Summary of the “Norwegian” regression results.....	41
Table 15: Summary of the “SAS” regression results.....	42
Table 16: Market shares of the airlines on domestic routes in Norway on 07.09.2020.....	53
Table 17: Market shares of the airlines on domestic routes in Norway on 11.09.2020.....	53

1. Introduction

In airline industry, as well as in many other industries, pricing has always been one of the most important processes. The price of the airline fare is determined by many different factors, and, thus, has ability to reflect strategy of the specific airline company.

This thesis focuses on the airline market in Norway with the purpose to explore the pricing strategies of two leading companies on Norwegian airline market, Scandinavian Airlines (SAS) and Norwegian Air Shuttle (Norwegian) for domestic flights. The chosen method is linear regression analysis of pricing data in statistical software RStudio.

The purpose of the statistical analysis contained in this thesis is to test the hypotheses that our datasets do not provide evidence that certain factors are correlated with the average airline fare price against the hypotheses that there is enough evidence of such correlation. For the cases with observed correlation, the goal is to find out if the correlation is positive or negative. It is therefore important for us to know if the change in specific factor affects the average ticket price and how.

The thesis focuses on such factors as duration of the flight, market share of the airline on the route, competition with regard to flight departure, time of the day when the flight departs and fare characteristics as on explanatory factors for change in ticket prices on average. The pricing datasets include pricing data for flights with departure on two different days of the week: Monday 07th September 2020 and Friday 11th September 2020, which are analysed separately. This also makes it possible to compare the analysis results for two different samples and see if the explanatory characteristics affect the ticket price differently depending on the weekday when the flight departs.

For both 07.09.2020 and 11.09.2020 samples, there conducted four types of analysis:

1. The general analysis which considers both observations for Norwegian and SAS in the analysis
2. The “economy subset” analysis which only considers the subset of well-comparable economy-class tickets of both Norwegian and SAS
3. The separate analysis of Norwegian pricing data

4. The separate analysis of SAS pricing data

What makes this research interesting is that it is conducted during the times of COVID-19 pandemic, which severely affected the airline market worldwide. There are factors that, according to economic literature, (insert link) are proven to be correlated with the fare price, such as market share and degree of competition. We are interested if our data supports that changes in such variables are correlated with changes in the average ticket price, while there are other random factors that may affect fare prices due to pandemic.

According to Norwegian Competition Authority (Norwegian: Konkurransetilsynet), as a result of the COVID-19 pandemic, “the aviation sector has been granted a temporary exemption from the Competition Act's ban on anti-competitive cooperation. The regulations entered into force on 18 March 2020 and will apply until 31 December 2020. Cooperation must not go beyond what is strictly necessary to ensure the maintenance of socially critical functions in connection with the COVID-19 pandemic, and must, as far as possible, ensure the efficient use of resources and the interests of customers. The exception is particularly aimed at making it possible to maintain passenger and freight transport in Norway, in order to ensure citizens access to necessary services” (Konkurransetilsynet 2020).

It is necessary to note that our data collection is done during the period without serious restrictions in Norway, more specific, from 08th August 2020 to 10th September 2020. However, the fact that the situation on the airline market is different due to pandemic should be kept in mind.

The thesis is structured as follows. The Introduction section is expanded and further on continues with relevant theory sub-section, which explains pricing processes on the airline market, and previous research sub-section, which highlights some papers relevant to our research. Section 2 includes general presentation of the airline market and its actors, as well as relevant knowledge about Norwegian airline market and two of its representatives, SAS and Norwegian. Section 3, firstly, contains theory on existing types of data, secondly, our method - linear regression and different estimation techniques. Finally, the necessary information about our data collection process is also provided in Section 3. In Section 4 we describe the method implementation and present results of our analysis. Section 5 consists of summary, concluding remarks and future research prospects.

1.1 Revenue management

The natural purpose for any firm is to maximize its profit. Considering the airline market, costs that are charged for the ticket, especially short-term costs, are originally fixed, and, due to that, profit maximization is deeply connected to revenue maximization in the context of airline industry (Malighetti, Paleari et al. 2009).

In other words, low variable costs and high fixed costs lead to an environment where maximization of revenue is almost equivalent to maximizing profits. Revenue management is a discipline that originated exactly in the airline industry in the 1970s following the deregulation of the US airline market (Strauss, Klein et al. 2018).

The definition of revenue management refers to the use of data and analytics to make forecasts about customer demand and then, based on performance, to choose the best management decisions, such as pricing and segment distribution, in order to optimize profits or revenues. Since this term has been established, it has been implemented in a variety of sectors such as transport, hotel business, advertising and many others. The distinction between revenue management and more general pricing types is that revenue management typically uses advanced IT systems that automatically process data in order to obtain forecasts for demand and to optimize management accordingly.

Due to fact that roots of revenue management are linked to the airline industry, revenue management is often associated with the regulation of availability.

The following is the central issue of the availability regulation. A goal of a business is to sell goods out of a finite product set before the certain date. The goods are to be sold to different consumers that arrive before this date but will use the services assigned to those goods afterwards. Air travel ticket is the natural example of such a good. For not a long period of time costs of such goods and capacity for the relevant services are typically fixed. A very common situation is that costs of such goods as well as capacity for the relevant services are decided in the short term. Furthermore, capacity is decreasing and sometimes correlated with high fixed costs, while marginal costs are often due to the sale of an additional unit of product and are more or less low.

Such a structure of costs allows revenues to be a substitute for profits. The described situation leads company to the goal of maximizing the overall revenue with help of revenue management. With regard to the products customers have different preferences and this means willingness-to-pay of those customers is also different. That is why the choices made by the customers will differ and depend on the offered products. The company, in its turn, should be able to manage the sales during the selling horizon in a right way, and availability control is responsible for this. Summarized, availability control means deciding which exactly products to offer at which time during the selling horizon.

In the airline industry, a product is commonly represented by a ticket for the desired route in a specific compartment of the aircraft, associated with a certain booking class which may be linked to some booking restrictions (Strauss, Klein et al. 2018).

1.2 Price discrimination

“Nearly all firms with market power attempt to engage in some type of price discrimination”(Varian 1989). Price discrimination is strongly connected to revenue management and is defined as setting different prices to different customer groups for the same good. Firms may charge prices guided by specific criteria, for instance, price for a flight may depend on time of departure, purchased quantity, age of customer and time of purchase. The common practice is to set higher prices for departures during the peak hours and make early purchase discounts. Price discrimination should not be confused with price dispersion, because the latter assumes several companies on the market who set different prices for the same type of good, however, price discrimination assumes it is one company/seller who does this.

The key principle of price discrimination is to use different price elasticities of demand. For instance, if we consider the airline market, some groups of air travellers have relatively more inelastic demand compared to others, and it means that they are willing to pay a higher price than others for a specific good. For this reason, firm charges higher prices for customers who are willing to pay more, thus increasing the profits. Those customers who have more sensitive reaction on change in prices and thus have relatively more elastic demand, are commonly interested in discounts and special offers. Because of this, such a separation of consumers into groups brings benefits to firms: it reduces consumer surpluses.

In real world situations, price discrimination may work in a little different way. Price discrimination often means setting a different price for a product that is slightly different. Air travellers, for example, receive more service in business cabin class than in economy class.

1.3 Price discrimination and the airline market

First type of price discrimination that is present on the airline market is so-called refund-based price discrimination. In terms of purpose of the trip, air travellers are often classified into two segments: leisure travellers and business travellers. While the former group has relatively low value of time, the latter, vice versa, value their time relatively more and, therefore, are likely to be interested in booking tickets for the specific day. Furthermore, business travellers commonly prefer fares which include refundability option in order to prevent the loss of money in case they will not have possibility to travel on this day. The listed facts give airline companies the incentive to set ticket prices as a function of refundability, in other words, enforce business travellers to pay higher price compared to leisure travellers for the specific flight compared to leisure travellers.

Another type of price discrimination that characterizes airline market is inter-temporal price discrimination, which is one of the forms of dynamic pricing. Dynamic pricing is a type of time-of-use pricing where prices can differ during the peak period on a limited number of days (Faruqui 2012).

Inter-temporal price discrimination in context of the airlines means that air companies set ticket price as a function of the time of the purchase. The core is that airlines, in order to make more profit, commonly use the strategy of classifying their customers into groups depending on the day when the customers want to book a specific seat.

1.4 Previous research

There were many studies that analysed impacts of different factors on airline fare prices. For instance, Malighetti, Palesari and Redondi in their paper made an analysis of pricing strategies of low-cost airline Ryanair (Malighetti, Palesari et al. 2009). Among their findings was that average price on the route increases with the following characteristics: length of the route, flight frequency on the route and fully booked flights percentage. It is also interesting that they found a negative correlation between price and Ryanair's offered seats share at the airports of origin and destination.

Pels and Rietveld in their paper based their research analysis on the evidence from London-Paris market and explored if the pricing behaviours of airline companies are correlated with pricing behaviour of their competitors (Pels and Rietveld 2004). Both strategies of network and low-cost carriers were explored, and the general finding is that network airlines do not follow the price adjustments of low-cost carriers, rather some airlines lower prices when there is an increase in prices of potential rivals. Moreover, in this paper they also found the evidence that all air companies follow the pattern of increasing prices as the day of flight departure gets closer.

In their paper Giaume and Guillou, who investigated European airline market via collecting the ticket price data of all flights from the Nice airport to European directions (Giaume and Guillou 2004), as well as Gaggero and Piga, who explored airline competition on routes between UK and Ireland (Gaggero and Piga 2010), and others also give the evidence that ticket prices are commonly increasing over time till day of departure. Alderighi et al. in their paper investigated pricing strategies on European airline market on example of London-Amsterdam route and provide the evidence that average prices increase by about 3% every day prior to departure, with an aggregate increase of 80% during the last twenty days before the departure (Alderighi, Cento et al. 2011). Escobari and Jindapon in their paper explore the refund-based price discrimination via investigating of the dataset of US airlines (Escobari and Jindapon 2014). The aim of the paper was to find out how the air companies in US use the function of refundability of the ticket on the monopoly routes to screen the customers who are not certain about their travel. The paper provides the evidence that opportunities of refund-based price discrimination through refund contracts decline closer to the departure date and travellers learn about their demand.

Clark and Vincent in their paper based their research on US airline market and found the evidence of airlines' dynamic pricing behaviour (Clark and Vincent 2012). It was also found that some carriers respond to their competitors pricing changes and load factors, and moreover, increase prices when there are no more places available on flights of rival companies.

2. The airline market

The airline industry market of today is constantly evolving. The aviation industry, about thirty years ago, was essentially a public service led by government agencies. The airline market today looks very different and the explanation for this is that prices and levels of service are determined by the customer.

The number of airline companies grew quite rapidly across all types of airlines. Consequently, fares in the airline industry became more competitive. With the elimination of numerous restrictions and rules, it became possible for customers to have more travel options and, not least, cheaper fares.

Industries usually select a business model that may be different from the models chosen by rivals, and airline industry is no exception here. Airlines are classified within the industry into the following three categories:

- Network carriers, often called legacy or full-service airlines
- Low cost carriers or budget airlines
- Charter airlines, holiday/leisure airlines that are owned by tour companies and focus on transportation of tourist groups to the destination

We will further on focus on the first two categories, since they are the most relevant to research contained in this thesis.

2.1 Network airlines

A network/legacy airline is an airline that focuses on providing its passengers a full-service flight. It assumes that such options as checked luggage, meals and beverages on board etc. are included in the ticket price, as well as stays in airline's own airport lounges. However, network airlines obviously offer various service classes, usually Business class, First class and Economy class.

Full-service carriers also tend to operate over a large domestic and international route network, and long-distance routes lead passengers via the main hub on to the destination. Network airlines are most often members of airline alliances. Moreover, it is very common for them to offer frequent flyer programs, in other words, loyalty programs that encourage frequent air travellers to use services of this specific airline or its partners.

Outstanding customer service is key for network airlines. Most of today's airlines fleets consist of wide-body and narrow-body jets that serve long, medium and short-range routes with aircraft size and capacity optimized to meet passengers demand.

The largest network airline in Europe as for 2019, according to the website Skycop.com, is Lufthansa, which services were used by 145,1 million passengers in 2019 (Diagilev 2020).

2.2 Low cost airlines

Low-cost airlines are the strong competitor for network airlines nowadays. According to Statista.com, low cost carriers in Europe had 33.1 percent of the total seat capacity in the region in 2019 (Mazareanu 2020). Moreover, the largest share of the whole European airline market, as for 2019, is taken by low-cost airline Ryanair, which transported 152.4 million passengers in that year.

The business model of low-cost carriers has evolved considerably over the last twenty years. The model focuses on business and operating strategies that minimize airline costs, and this is the biggest advantage over the network airlines, as the operating expenses policies of low-cost airlines are considerably smaller.

The market share of low-cost airlines in Europe grew by 10 percent in 2019 compared to the level of 2009 (Mazareanu 2020), however, they expectedly faced a competitive response from network carriers during this period. Network airlines clearly needed to retain their position on the market and, consequently, began offering alternative low-cost flight options, in other words, selling budget, economy-class tickets. Nevertheless, as network airlines get into competition within the low-cost branches, they still face difficulties in implementing such a model as efficiently as low-cost airlines.

Many cabin crew service offers are usually not included in a regular ticket of a low-cost airline and charged as additional costs to certain passengers depending on their demand for it. Such additional fees shall be considered as sources of potential additional revenue for low-cost carriers.

All in all, the business model of low-cost airlines has stimulated demand for traveling, as well as appeared to be efficient all over the world and has increased the growth of air transportation.

2.3 Airline market in Norway

Norway is located on the edge of Europe, has an open economy and extensive economic relations with other countries. This often presupposes good air connectivity and convenient air transport services to the rest of the world. In 2018, Norway was the European country with the highest number of air passengers to, from and in the country in terms of population size. Indeed, many of the travellers were visitors from another countries, nevertheless, Norwegians are among the nations who travel the most by air in Europe. This fact can be explained by an open economy and high income, as well as by dispersed settlement. In view of the fact that Norway is an elongated country with a challenging topography, good access to transport is necessary if citizens are to be able to travel within the country, to reach Oslo and other major cities, important institutions such as hospitals, public authorities and other infrastructure. There are several parts of Norway where flights are the only practical choice for more or less long journeys, and that is why good flight offers are needed. Good accessibility has an effect on the labour and the education market, since it makes it realistic to live in one location and work or study in another city (Regjeringen.no 2019).

The largest part of Norwegian airport network is owned and operated by Avinor AS where all the shares are owned by the Ministry of Transport and Communications (Norwegian: Samferdselsdepartementet). The company's responsibility is to operate and develop a domestic network of airports and air security services for civil aviation. The services of aviation security are arranged by the subsidiary Avinor Flysikring AS. Avinor must operate its business in an efficient, safe and environmentally friendly manner, as well as ensure good accessibility for all groups of travellers (Regjeringen.no 2019).

The domestic airline market in Norway was recently dominated by three companies: Scandinavian Airlines (SAS), Norwegian and Widerøe. Together they accounted for 99 percent of domestic traffic. There is competition between SAS and Norwegian on most of the major routes in the Norwegian market. SAS and Norwegian are the big companies on the international market, with 27 and 35 percent of the offer, respectively. Widerøe's route network mostly represents shorter flights from and to local and regional airports along the coast from Oslo to Kirkenes (Regjeringen.no 2019). The Hungarian low-cost company Wizz Air entered the Norwegian domestic airline market in October 2020 and caused the decreased levels of prices

(Vege 2020), however, it is not considered in this thesis due to the fact that the dataset that supports the thesis conclusions was collected before October 2020. Levels of competition in Norway are regulated by Norwegian Competition Authority.

In this thesis we will focus on main routes in Norwegian domestic airline market and, therefore, its two well-established dominant players, SAS and Norwegian.

2.4 Scandinavian Airlines

Scandinavian Airlines (SAS or Scandinavian Airlines System) is a multinational network airline in Norway, Sweden and Denmark and in the same time a leading Scandinavian carrier. The headquarters of the company are based in Stockholm, Sweden. The main hubs of SAS-airlines are Stockholm-Arlanda, Copenhagen and Oslo Gardermoen airports. SAS was founded on August 1, 1946 and was as a result of a collaboration agreement between Swedish Intercontinental Airlines (Svensk Interkontinental Lufttrafik AB), Danish Air Lines (Det Danske Luftfartselskab A/S) and Norwegian Air Lines (Det Norske Luftfartselskap AS, to not be confused with Norwegian Air Shuttle low-cost airlines). This agreement was made with the goal to gather all resources in order to improve the intercontinental air traffic to and from Scandinavia.

In 1997 SAS founded Star Alliance, which is the oldest, largest and the most representative alliance in aviation today. It was established together with Lufthansa, United Airlines, Air Canada and Thai Airways.

In 2002, SAS gained a monopoly on the domestic market in Norway through buying Braathens airline, which was in debt. However, monopoly did not last for a long time, because in 2003 Norwegian Air Shuttle established itself as an independent airline. After that, SAS divided its business into four independent companies within the SAS group: SAS Sweden, SAS Denmark, SAS Braathens and SAS International. SAS Braathens changed its name to SAS Norway in 2007 (sas.com).

SAS is a full-service network company which means that SAS offers number of services for its customers both in the air and on the ground in addition to the air transport itself. Therefore, the company has extensive cooperation with other airlines and operates with different classes on board.

As for 2018, SAS had nearly 30 million passengers and managed 222 thousand departures with 157 aircraft in service. According to the SAS website, the vision of SAS is to make life easier for frequent travellers in Scandinavia. As consequence, the target community for SAS is frequent travellers inside and from Scandinavia. One of the strategic focus areas for SAS is to win Scandinavia's people who travel frequently by offering an attractive seasonal schedule and network. Moreover, the priority is to continue evolving respected service concepts and to deliver productive digital solutions along the travel chain. These are the forms SAS follows to increase its competitiveness (sas.com).

2.5 Norwegian Air Shuttle

Norwegian Air Shuttle (often referred to as simply Norwegian) is a low-cost carrier with headquarters in Fornebu, Oslo. The company is the second largest after the Scandinavian SAS airline and the third largest fleet, and low-cost airline in Europe (Diagilev 2020), after Ryanair and EasyJet.

Norwegian airlines were founded in 1993. Initially they started flights on the west coast of Norway with Fokker 50 airliners, in cooperation with Braathens airlines. This collaboration was in power until 2002, when Norwegian launched domestic routes in Norway with Boeing 737-300 aircrafts. In 2003 Norwegian was listed on the Oslo Stock Exchange and 2005 was the first year when Norwegian received profit (norwegian.no).

“Everyone should afford to fly” is the slogan of Norwegian airlines. It reflects their vision, principles, organizational goals and objectives. According to the Norwegian Air Shuttle website, one of Norwegian's main goals is to attract new customers by providing high-quality travel experience, affordable pricing and freedom of choice. Through freedom of choice, Norwegian assumes more reasonable offers for price-conscious consumers and a more comprehensive package for those who want additional service. Norwegian focuses on offering a wide variety of tourist destinations as well as an extensive network of common routes to major airports for business travellers. Moreover, the aim of Norwegian is to continue developing high-quality and cost-effective goods and services (norwegian.no).

In 2019 Norwegian becomes "World Best Low-cost Airlines for Long-haul Flights" by Skytrax World Airline Awards for the fifth consecutive year by Skytrax (norwegian.no).

3. Method and data

The collection of data plays an important role in every statistical analysis. In order to carry out the study, various researches gather information in number of ways, depending on the purpose of the further analysis. In terms of analysis there exist two types of data: cross-sectional and time-series data. Cross-sectional data refers to observations on multiple units at a single time point. It is believed that units in this case are, for example, people, companies or countries. Time series, on the other hand, refer to observation of only one unit, but chronologically over time, for instance, a day, a week, a month. Panel data, in its turn collects findings from multiple units at different times and, therefore, is the combination of cross-sectional and time-series data. In terms of panel data, it is common to use the subscript i to refer to units and t to refer to time. For instance, observation Y_{it} is collected for unit $i = 1, 2, \dots, N$ over all periods of time $t = 1, 2, \dots, T$ and, as a result, the dataset is of $N \times T$ size.

Among the advantages of panel data are its ability to observe both individual and common group behaviors and to measure certain additional statistical effects. Panel data is thus considered to be more insightful compared to pure cross-sectional or time-series data. In addition, the collection of individual groups in a single time series can lead to estimation biases, although panel data helps to solve this problem by minimizing biases.

Panel datasets can also be classified into two types: balanced and unbalanced panel data.

Balanced panel data has the same number of observations for every unit, unlike unbalanced panel data, which assumes some incomplete cells or missing values in dataset due to a lack of observations at some point in time for some of the units.

3.1 Linear regression analysis

In statistics, the common method for analyzing panel data is regression analysis. The aim of such analysis is to estimate the characteristics of the population by using a random sample taken from that population. A regression model defines correlation between various data points, and its purpose is to determine the relationship between a dependent variable and one or more independent variables which are also referred as explanatory variables.

If we believe that relationship between the explanatory variables and the dependent variable is linear, it would be natural to construct a linear regression model. Estimation analysis of such

model is conducted via Ordinary Least Squares (OLS) approach which minimizes the sum of squares in the differences between the dataset-observed and function-predicted values of dependent variable.

A simple linear-regression equation can be written as follows:

$$y_{it} = \beta_0 + \beta x_{it} + \varepsilon_{it} \quad (3.1.1)$$

The left side of the equation, y_{it} , stands for the dependent variable while the right side is the explanatory part. β_0 and β are coefficients which need to be estimated, where β_0 is the intercept coefficient and β is the slope coefficient belonging to independent variable x_{it} . In this example, there is only one explanatory variable, however, in practice it is typical to include two or more variables of this kind, and the purpose of OLS thereafter is to estimate all β coefficients in the equation. In this case it is multiple regression model.

Index i stands for individual/group of individuals and index t denotes time. The equation also includes an error term ε_{it} which contains the variation in dependent variable that cannot be clarified by independent variables. The sample estimate of the error for every observation is shown by residuals that indicate the difference between the observed value and the value fitted by the model.

The core of regression models is that they help us to find the statistical evidence of positive or negative relationship between specific explanatory variables and dependent variable. The process assumes testing the null hypothesis (H_0), the meaning of which in statistics is that there is no correlation between the specific independent variable and dependent variable. The alternative hypothesis (H_1) is that the correlation at some degree is present. If result of the test shows that null hypothesis must be rejected, we accept the alternative hypothesis, and it is said that the result is statistically significant. In this case we conclude that data sample provides support to correlation between two variables. We can therefore proceed with interpretation of the correlation coefficient. The common approach to identify if we should reject null hypothesis is to examine p-value, which shows how high is the probability of getting this specific result of test statistic or greater if the null hypothesis is true. If the p-value is very small, under significance level value, it means that such a probability is very low, so the null hypothesis is rejected. Studies often choose number 0,05 as significance level, however, some prefer to choose 0,1.

Again, it is common to use linear regression for panel data analysis. However, in order to ensure that the OLS method produces reliable and best possible estimates, one should make sure that the regression model satisfies the following six assumptions:

1. Linearity assumption

The name Linear regression makes it obvious that the appropriate model must be linear in all parameters, both in coefficients which can also be multiplied by a random variable, and in error term.

2. The average value of the error term is equal zero

The average value of the error term is assumed to be zero, otherwise some part of the error will already be predictable. This would be a contradiction since in this case, the logic suggests us that this component must be included in the regression model itself rather than remain in the error term. This situation will lead to a systematic error which is associated as coefficient estimates bias. In other words, the model would be inaccurate, since it would consistently underestimate the observed values.

3. No endogeneity in the model

Part of the variation in the dependent variable is explained by the error term, but in case there is endogeneity, some independent variables will also be correlated with error and can be used to predict it. This is a contradiction due to fact that error term must only contain unpredictable random error. Such situation, as well as breach of the Assumption 2, will lead to bias in estimates, because OLS method will assign more variance to independent variable than it should, while this excess variance actually belongs to the error term.

4. No heteroskedasticity in the model

It can happen that the linear model does not meet assumptions of homoskedasticity and in this case OLS regression method may considerably lose its efficiency (Breusch and Pagan 1979). It is initially assumed by OLS regression that the residuals of the error term are homoscedastic, in other words, that they are continuously spread across the plot, since the variance in the population is constant. Heteroskedasticity, in its turn, is an unequal spread of residuals. This issue can occur if, for instance, we exclude some variables from the model or if there are some observation values in the dataset that are too small or too large compared to the rest of observations. Heteroskedasticity makes estimates less precise, and, as a result, it can cause the

false conclusion that some parameters in the model are statistically significant when they are not. It happens because the variance which OLS technique uses to calculate the statistical values is underestimated, and problem of heteroscedasticity is not taken into account by OLS.

5. No autocorrelation in the model

It is assumed by OLS that observations in the error term are not correlated with one another, and if this is not the case, it is said that the model exhibits autocorrelation or serial correlation. The problem of autocorrelation moves estimate further away from correct values. It typically occurs in time-series data and in practice it means that observations in the series from the certain period can be predicted by the values from the previous period. Pure cross-sectional models may also exhibit autocorrelation if observations are correlated with each other in some other way, for instance, if people from one observed group are more likely to provide similar answers than people from other groups, because of the certain group-characteristics (e.g. group of students). When we analyze panel data which is the combination of cross-section and time-series, we should naturally always control for serial correlation.

6. No multicollinearity in the model

Independent variables in a regression analysis must be uncorrelated with each other, otherwise the model will be prone to multicollinearity.

When all other independent variables excluding the one currently considered independent variable are kept constant, the mean change in the dependent variable for every 1-unit change in the considered independent variable is shown by the regression coefficient. On the other hand, in case of multicollinearity, another independent variable is correlated with the considered variable, variations in first variable are also associated with changes in the second one. The higher is the degree of this correlation, the harder it becomes to change one explanatory variable when keeping another constant. Thus, such a strong correlation can lead to difficulties in fitting the model and interpreting the findings, while the estimate coefficients are further away from their actual values.

3.2 Pooled OLS, Fixed Effects and Random Effects models

Panel data regression models can be classified into two types: homogenous and heterogeneous models. By homogeneous panel-data model one assumes that the parameters in the model are common across the units, while in heterogeneous panel data model the parameters are expected to vary.

Considering a simple linear regression model, if constant β_0 is identical for all groups of units at any time, and coefficient β does not vary across all groups and time, we say that the model is homogenous. This means that only error term is responsible for group variety. However, some models may also include the specific constant for each group of units, which is the sign of heterogeneity.

The model which includes only the observable characteristics, for instance, age, height, size is called Pooled Ordinary Least Squares model (Pooled OLS), and it considers all panel data as one pooled dataset. The term α in equation is constant and both β and c are assumed to be estimated by Pooled OLS technique.

$$y_{it} = \beta_0 + \beta x_{it} + c + \varepsilon_{it} \quad (3.2.1)$$

Heterogeneity, in its turn, is connected to some non-observable effects in the model. For example, if model units are individuals, personal characteristics or skills of those individuals are likely unobservable. In such situations, instead of Pooled OLS, it is common to choose either Fixed Effects or Random Effects panel data model.

A Fixed Effects panel data model contains unit- or time-specific characteristics which are assumed to be correlated with the included observed characteristics. Equation gives us the example of such model with the unit-group specific intercept, c_i . The key point is that term c_i consists both of constant effect and some unobservable but important fixed effect which one should account for while performing the estimation. If one attempts to apply Pooled OLS to model with fixed effects, it will not give consistent results due to fact the unobservable effect will not be estimated. Fixed Effects technique estimates β and c_i for every group in the panel dataset:

$$y_{it} = \beta_0 + \beta x_{it} + c_i + \varepsilon_{it} \quad (3.2.2)$$

A Random Effects model, in turn, also assumes unit- or time-specific characteristics, however, they are not correlated with observed effects. Random Effects model includes effect which does not have a true value, but rather randomly distributed, and one should estimate the parameters of this distribution. Random effects in a model affect its error term, and Pooled OLS will not appear to be efficient method to estimate such model. Random Effects model can be represented by the following equation:

$$y_{it} = \beta_0 + \beta x_{it} + (c_i + \varepsilon_{it}) \quad (3.2.3)$$

The practical question that arises is how to choose the regression model in order to analyze the data in the most appropriate way.

In practice, in order to choose the type of the OLS regression model that will lead to the most appropriate analysis of the specific data, it is natural to conduct statistical tests. Chow test is the statistical test which, based on its results, allows us to determine which model would be more appropriate for estimation of the given panel data, Pooled OLS or Fixed Effects model. Hausman test, in turn, is performed in order to make choice between Fixed Effects and Random Effects models. Finally, Lagrange Multiplier (LM) test is aimed to determine if Pooled OLS or Random Effects model is the best to use in specific situation. The core of performing the statistical tests is to formulate the null hypotheses appropriately and test it. Table 1 summarizes the criteria for the correct choice of the model, which depend on the p-value produced by the test.

Table 1: Choice of the model based on results of statistical tests

	Chow test	Hausman test	Lagrange Multiplier test
Accept H0, p-value > 0,05	Pooled OLS	Random Effects	Pooled OLS
Reject H0, p-value < 0,05	Fixed Effects	Fixed Effects	Random Effects

When the same unit sample is observed over time, Fixed Effects and Random Effects models are those models that will appear to be efficient for analysis (Wooldridge 2019).

3.3 GLS estimation

In some cases, OLS estimation will not give the best results regardless what type of model is employed, because the model violates basic OLS assumptions. It can be challenging to simply say that model is prone to serial correlation or heteroscedasticity. However, if such issues are not

resolved, the best possible estimates will not be given by OLS technique. In order to detect problems of this type, several tests can be performed. For instance, Breusch-Pagan test is aimed to test the model for heteroscedasticity, while Durbin-Watson test is employed to detect autocorrelation.

If due to the dataset structure, residuals of the error term in the linear model are not homoscedastic and, moreover, the model exhibit serial correlation, the possible solution is to use Generalized Least Squares (GLS) estimation method, which takes such features of the dataset into account.

3.4 Data collection process

The general aim of this thesis is to determine if there is evidence from the Norwegian airline market that specific factors and characteristics affect the ticket price. In order to do so, as well as to see if the general results are consistent with economic literature on airline pricing strategies, it is natural to obtain the necessary data on fare prices. Two air companies on the Norwegian airline market are considered in this study, namely SAS and Norwegian. In order to obtain pricing data for flights operated by those air companies, we used the official SAS and Norwegian websites and collected the data manually. It can be related as primary data collection, because our datasets are first-hand data obtained by the one conducting the research.

For the analysis, it was decided to consider only direct domestic flights in Norway. Therefore, flights with stopovers are not included in the dataset. It was initially expected that the data for the most popular international flights from the most trafficked Norwegian airport Oslo Gardermoen would be collected and included in the dataset, but the situation on the airline market was complicated by the COVID-19 pandemic. Due to higher uncertainty connected to international travels, it was therefore agreed to focus only on domestic flights. Table 2 provides the evidence that number of international flights departing from Norwegian airports is decreased dramatically (-62.3 %) compared to levels of 2019.

Table 2: Flight movements in Norway – departure and arrivals

	September			Year to Date		
	2020	2019	Change	2020	2019	Change
Domestic	30,530	40,391	-24.4%	236,181	330,664	-28.6%
Scheduled	29,324	39,343	-25.5%	225,684	320,951	-29.7%
Charter	456	405	12.6%	4,249	4,785	-11.2%
Freight	750	643	16.6%	6,248	4,928	26.8%
International	4,890	17,438	-72.0%	55,972	148,301	-62.3%
Scheduled	4,004	15,514	-74.2%	47,493	131,122	-63.8%
Charter	257	1,382	-81.4%	3,669	12,228	-70.0%
Freight	629	542	16.1%	4,810	4,951	-2.8%
Offshore	3,327	3,582	-7.1%	31,383	31,417	-0.1%
Sum movements	38,747	61,411	-36.9%	323,536	510,382	-36.6%
Other civil flights	9,470	9,331	1.5%	70,759	82,225	-13.9%
Sum all categories	48,217	70,742	-31.8%	394,295	592,607	-33.5%

Source: Avinor.no, Monthly report (Avinor.no 2020)

All flights included in the dataset are run by either SAS or Norwegian Air Shuttle and Norwegian Air Sweden. The latter is an integrated subsidiary of Norwegian Air Shuttle, which operates number of flights in Norway.

When choosing the dates, it was crucial that prior data collection takes place during the time when random shocks which would have a major effect on the dataset are minimized. It was originally decided to collect data during March 2020; however, due to the COVID-19 pandemic and many flight cancelations in March and April, it became difficult to obtain a credible dataset. The data collection period therefore had to be postponed until August and September when the domestic flight situation became more stable.

The choice of flights for the dataset was based on the SSB statistics (Statistics Norway, Norwegian: Statistisk sentralbyrå) (SSB.no 2019), and it was decided to collect data for flights between the five most trafficked airports in Norway as for the fourth quarter of 2019: Oslo Gardermoen, Bergen Flesland, Trondheim Værnes, Stavanger Sola and Tromsø Langnes (see tables 3 and 4). With support of historical information, we believe that the airports with the highest traffic in 2019 will also have the highest traffic during the relatively stable period in 2020.

Table 3: Airports in Norway with the highest traffic, 4th quarter 2019

Airport	Number of passengers	
	Arrival	Departure
Oslo Gardermoen	1 479 211	1 452 023
Bergen Flesland	524 739	533 684
Trondheim Vaernes	445 000	457 780
Stavanger Sola	349 984	352 362
Tromsø Langnes	277 517	276 768

Source: SSB.no (SSB.no 2019)

Table 4: Traffic between the chosen Norwegian airports, 4th quarter 2019

Departure from	To				
	Oslo	Stavanger	Bergen	Trondheim	Tromsø
Oslo		214 533	250 988	269 890	152 849
Stavanger	218 958		77 521	14 967	0
Bergen	259 597	76 854		53 795	13 633
Trondheim	278 423	17 279	55 006		2 157
Tromsø	152 493	0	13 123	1 501	

Source: SSB.no (SSB.no 2019)

For storing the data, it was decided to use Excel spreadsheets. The data collection technique involved the collection of prices for all flights between selected airports that depart on certain days, during 30 days prior to departure. In order to improve the reliability of the dataset, we collected data twice a day. First collection period was morning-afternoon, between 11PM and 13PM, and the second one was happening at late evening, between 10PM and 12AM. The hours were selected in such a way that it would be feasible and convenient for the data collection to be carried out in those hours for a period of thirty days. In order to make sure that price information is reliable and not affected by external factors (e.g. cookies), the measures such as private network were implemented.

We obtained four datasets and we included all ticket prices for both SAS and Norwegian flights leaving on 7th and 11th September 2020. The 7th September is chosen as a random weekday, Monday, while the 11th of September is Friday, which is a common day for office workers to go on a domestic weekend trip to another city without taking days off job, considering that office workers in Norway have a five-day work week.

Ticket-price data were collected for all adult tickets provided by SAS and Norwegian, so fare prices for children and youth discount tickets are not included in the dataset. Both SAS and Norwegian sell several types of adult fares so that customers can decide which type of ticket fits their preferences more.

The types of tickets available on the SAS website for domestic trips fall into the following two categories: SAS Go and SAS Plus. The former can be referred to as economy class tickets, while the latter category assumes more costly fares that offer customer more service. SAS Plus tickets are available in three sub-categories: SAS Plus Smart, SAS Plus Pro and SAS Plus Full Flex, sorted from the most affordable to the most expensive.

Norwegian, in turn, offers three types of adult fares: LowFare, LowFare+ and Flex, sorted from the cheapest to the most expensive. The first two categories can be related to the economy class, while Flex offers more flexibility and service. It is clearly noticeable that Norwegian, as a low-cost airline, provides less variety of costly fares than the network carrier SAS.

Besides the facilities included in any ticket offered by SAS or Norwegian, additional attributes such as luggage, wi-fi on board or a seat reservation may usually also be bought (see table 5 and table 6).

According to economic theory, our datasets are referred as panel data, because each of our datasets consists of 60 cross-sections which are collected during a time period of 30 days. This panel data is also referred as unbalanced panel data, because some variables in the dataset are missing.

Table 5: Overview of ticket types and service available for domestic travels

	SAS Go	SAS Plus		
	SAS Go Smart	SAS Plus Smart	SAS Plus Pro	SAS Plus Full Flex
Carry on 1×8kg	✓	✓	✓	✓
Checked baggage 1×23kg	✓	✓ ×2	✓ ×2	✓ ×2
Rebooking	For a fee	For a fee	✓ + paying price difference	✓
Refundable	X	For a fee	✓	✓
Seat reservation	For a fee	✓	✓	✓
Fast Track & Lounge access	X	✓	✓	✓
Priority boarding	X	✓	✓	✓
Wi-fi on board	For a fee	✓	✓	✓

Source: SAS.com (sas.com 2020)

Table 6: Overview of ticket types and service available for domestic travels

	LowFare	LowFare+	Flex
Carry on 1×10kg	✓ under the seat only	✓	✓ 1×15kg
Checked baggage 1×23kg	For a fee	✓	✓ ×2
Rebooking	For a 650NOK fee + price difference	For a 650NOK fee + price difference	✓
Refundable	X	X	✓
Seat reservation	For a fee	✓	✓
Fast Track at selected airports	For a fee	For a fee	✓
Priority boarding	For a fee	For a fee	✓
Wi-fi on board	✓	✓	✓

Source: Norwegian.com (norwegian.com 2020)

4. Method implementation and results

4.1 Choice of variables

As it was mentioned earlier in this thesis, regression model is the natural way to analyze how some factors, referred as explanatory variables, impact a specific variable. In particular, the model that explores if, how and in which extent price of good or service is affected by another variables is called a Hedonic model. In order to implement our analysis, it was decided to construct several models of this type and test them in statistical software RStudio.

In order to specify a regression model, it is necessary to choose the dependent variable as well as independent variables which are believed to explain the variation in this dependent variable. In our case, the dependent variable is price of the ticket, therefore it is important to consider several factors that may affect the price and set them as explanatory variables.

As it was mentioned, we have collected prices for several categories of tickets, therefore datasets provide price observations of four different categories for each SAS flight and three for each Norwegian flight.

The additional variable which have been collected and which could potentially be referred as explanatory variable in our regression model indicates how many seats were left for the current ticket price. However, SAS website only displays the remaining number of tickets if it is nine or less, and Norwegian if it is three or less. This resulted to many NA-variables in our datasets, which appeared to cause difficulties while analyzing the data in RStudio. Due to that, we decided to exclude this variable from our list of independent variables.

It would be natural to add variable which shows either how long does the flight lasts or how many kilometers is between departure point and destination point. We decided to create the variable *duration* which shows the length of flight in minutes. Another important explanatory variable is *time_to_departure* which shows how many days are left till the flight departs. This variable can be useful in order to find the evidence of inter-temporal price discrimination.

Another variable that may be influential on price is the market share of the airline on the specific route on this day. We have calculated market shares for SAS and Norwegian on each observed route for both 7th September and 11th September (see Table 16 and Table 17 in Appendix) by dividing number of flights of specific carrier on specific route this day by total number of flights

available on this route this day. It would therefore be possible to find out if market share of the airline on the route affects the average price of the ticket.

Our analysis also assumes presence of categorical variables in the model. In statistics, a categorical or dummy variable is an indicator variable that distinguishes an experimental group from a comparison group. In context of our regression analysis, such a variable is equal 1 when the observed individual or unit is in the experimental group and zero otherwise.

Firstly, we decided to add the dummy variables constructed based on the time of the day when the flight departs. Variable *early* takes value 1 if the flight departure is in the early morning, before 8AM, and 0 otherwise. This variable was introduced following the logic that some air travellers would prefer to book a very early flight so that they can be at the destination as early as possible in order to either take a connection flight or spend the whole day at the flight destination and take a flight back at the evening of the same day. We are therefore interested if there is evidence that air companies adjust prices for flights with very early departures. Following the similar logic, we have also introduced variable *late* which takes value 1 if the flight departs after 8PM and 0 otherwise. Another variable in this group is *early_evening*, which was generated to distinguish between the flights which depart after 4PM and earlier flights. The logic behind introducing this variable is to see if it makes different impact on average ticket prices depending on the day of flight departure. We assume that typical adult customer often has a five-day work week, and, therefore, may be interested to fly on a domestic weekend trip right after his Friday shift. Time point 4PM is chosen as an approximate benchmark. We are, therefore, interested if we could find such evidence that ticket prices were on average more expensive on Friday, 11th September, after 4PM compared to earlier departures on the same day, and if there is an identical pattern for Monday, 07th September.

Secondly, we considered to include the dummy variable which is based on data registration details. Variable *late_reg* therefore distinguishes among times of the day when our data was collected. The logic behind choice of this variable was our interest to investigate if ticket prices are on average more expensive or cheap if customer books ticket in the evening, or if there is no evidence of such connection. *late_reg* takes value 1 if the collection was conducted during the late evening (10PM to 12AM) and 0 otherwise.

Thirdly, we introduce dummy variables that are based on ticket characteristics, in order to achieve better explanation of the variation in prices. For our analysis we did not choose the variable which distinguish between economy ticket class and more costly fares that include refundability option, however, we choose several categorical variables that take value 1 if ticket category falls under the specific ticket characteristic or 0 otherwise. The logic behind this is that SAS provides not only economy class tickets SAS Go Smart and the most expensive fare SAS Plus Full Flex, but also the intermediate options SAS Plus Smart and SAS Plus Pro which have access to priority services. It may be obvious that SAS Full Flex and Norwegian Flex are the most expensive ticket categories, however it is important to include those variables in the model, due to their clear impact on fare price.

The dummy variables of this type, that potentially can be included in the regression model, are *lug_1* and *lug_2* which define if one or two units of luggage are included in the specific fare, *reb_1* and *reb_2* which indicate two different rebooking options, *ref_1* and *ref_2* which are indicators of two different refundability options, *seat* which takes value 1 if free seat reservation is included and *priority_1* and *priority_2* which define the availability of priority boarding options. However, according to one of the OLS assumptions, the model should not be affected by multicollinearity. In other words, if the model is too much specified, and there is a perfect correlation between some of the variables, it will cause difficulties for our analysis due to problems with fitting of our regression model. This situation will cause misleading results. This should be kept in mind while choosing how many variables should be added to regression model.

Finally, we add the dummy variable that show if the specific flight is in the competition with flights provided by another airline company. The variable *competition_departure* takes value 1 if there is a flight provided by the rival airline in the time range of 30 minutes prior to departure or after the departure. If this is not the case, the variable takes value 0. The inspiration for creating this variable was taken from the master thesis (Halvorsen and Skaug 2016), where authors analyzed SAS EuroBonus loyalty program. They used multiple linear regression in order to determine how specific factors affect the number of SAS EuroBonus points which customer earns on flight. The number of Eurobonus points was also associated with fare price. They considered rival flight as competitive, if it departs within one hour before or after the departure of the considered flight. However, their research focused not only on domestic market, but also

included international routes. We believe that most of flights on the domestic routes between the chosen airports are more frequent than international flights from Oslo Gardermoen. Therefore, it was decided to consider 30-minutes range for our analysis.

Table 7 collects the information about all variables which can potentially be included in our regression models.

Table 7: Suggested variables for regression model

Dependent variable	Ticket price	<i>TICKET_PRICE</i>
Explanatory variables	Time to departure	<i>TIME_TO_DEPARTURE</i>
	Flight duration	<i>DURATION</i>
	Market share	<i>MARKET_SHARE</i>
Explanatory dummy variables, 1 = true, 0 = false	Flight and competition	
	Availability of flight of another airline company in the time range of 30 minutes before or after the departure	<i>COMPETITION_DEPARTURE</i>
	Early departure (before 8AM)	<i>early</i>
	Late departure (after 8PM)	<i>late</i>
	Evening departure (after 4PM)	<i>early_evening</i>
	Registration details	
	Data was registered at the late evening (10PM-12AM)	<i>late_reg</i>
	Ticket characteristics	
	Possible to rebook without fee, but with fare difference	<i>reb_1</i>
	At least one unit of luggage is included	<i>lug_1</i>
	Two units of luggage are included	<i>lug_2</i>
	Possible to rebook without a fee	<i>reb_2</i>
	Refundable, but for a fee only	<i>ref_1</i>
	Fully refundable	<i>ref_2</i>
	Free seat reservation option is included in the price	<i>seat</i>
	Priority boarding is available for a fee	<i>priority_1</i>
Priority boarding is included in the price	<i>priority_2</i>	

In order to efficiently analyse average impacts of explanatory variables on average airline fare price, it would be convenient to merge Norwegian and SAS datasets which contain price observation for the flights departing on the same day. Due to challenges with data wrangling, it was therefore decided to not merge all four datasets together, rather work with merged SAS and

Norwegian datasets which contain price data for flights with departure on 7th September 2020 and 11th September 2020 separately. We thus intend to construct models, which we will name “07.09” and “11.09” further on in the thesis. These models have the same specification but use data from two different datasets. The results produced by two models will be compared after the analysis is conducted.

4.2 The general analysis

The following regression model is the general model which includes all proposed independent variables for our analysis. As it has already been mentioned, we will apply such model to two different merged datasets.

$$\begin{aligned}
 \text{ticket_price} = & \beta_0 + \beta_1 \text{duration} + \beta_2 \text{time_to_departure} + \beta_3 \text{market_share} + \beta_4 \\
 & \text{competition_departure} + \beta_5 \text{early} + \beta_6 \text{late} + \beta_7 \text{late_reg} + \beta_8 \text{lug_1} + \beta_9 \text{lug_2} + \beta_{10} \\
 & \text{reb_1} + \beta_{11} \text{reb_2} + \beta_{12} \text{ref_1} + \beta_{13} \text{ref_2} + \beta_{14} \text{seat} + \beta_{15} \text{priority_1} + \beta_{16} \text{priority_1}
 \end{aligned}$$

(4.2.1)

In order to receive the overview of in which extent dummy variables are correlated with each other, we construct the correlation matrix.

The matrix shows that the perfect correlation is observed between the variables *lug_2* and *priority_2* (Correlation value is equal to 1). In practice it means that tickets of both Norwegian and SAS follow the same pattern in *lug_2* and *priority_2*, in other words, all fares which include 2 units of luggage also include free priority boarding. For our regression model it means that we should exclude one of those variables. We decide to only leave the characteristic *lug_2*.

However, if we only remove *priority_2* and leave the rest of the dummy variables, the model still faces a singularity problem in RStudio, and some coefficients are deleted from the summary statistic due to that. Therefore, we decide to also exclude some other variables from the model. All in all, the generating and including certain dummy variables in the model will give ambiguous results due to incorrect model specification. For instance, the variable *priority_1* which shows that the ticket assumes that a fee should be paid for priority boarding service, will simultaneously compare such a ticket to both first class categories which assume that this service is included in price and SAS Go which assumes no priority boarding at all.

Due to possible difficulties with interpretation of the model coefficients we decided to exclude ambiguous ticket-restriction dummy variables and only leave variable *reb_2* which takes value 1 for Norwegian Flex and SAS Full Flex only, variable *lug_2* which distinguishes economy class and other tickets well because 2 units of luggage are only included in first-class fares. We believe that such a characteristic as luggage is important for customers when they book a ticket, however, we further on decide to exclude variable *lug_1*, because it does not distinguish between economy-class and first-class fares. Also, SAS does not provide such a ticket that includes only hand-luggage for domestic travels, so this variable would only be relevant to Norwegian fares. We conclude that our model will appear enough specified when it includes only three ticket-restriction dummy variables.

Table 8 provides the overview of variables that will be included in our regression analysis.

Table 8: Variables overview

Dependent variable	Ticket price	<i>TICKET_PRICE</i>
Explanatory variables	Time to departure	<i>TIME_TO_DEPARTURE</i>
	Flight duration	<i>DURATION</i>
	Market share	<i>MARKET_SHARE</i>
Explanatory dummy variables, 1 = true, 0 = false	Flight and competition	
	Availability of flight of another airline company in the time range of 30 minutes before or after the departure	<i>COMPETITION_DEPARTURE</i>
	Early departure (before 8AM)	<i>early</i>
	Late departure (after 8PM)	<i>late</i>
	Evening departure (after 4PM)	<i>early_evening</i>
	Registration details	
	Data was registered at the late evening (10PM-12AM)	<i>late_reg</i>
	Ticket characteristics	
	Two units of luggage are included	<i>lug_2</i>
	Possible to rebook without a fee	<i>reb_2</i>
	Fully refundable	<i>ref_2</i>

The model that we will test further on is the following:

$$\begin{aligned}
 \text{ticket_price} = & \beta_0 + \beta_1 \text{duration} + \beta_2 \text{time_to_departure} + \beta_3 \text{market_share} + \beta_4 \\
 & \text{competition_departure} + \beta_5 \text{early} + \beta_6 \text{late} + \beta_7 \text{late_reg} + \beta_8 \text{lug_2} + \beta_9 \text{reb_2} + \beta_{10} \text{ref_2}
 \end{aligned}$$

(4.2.2)

The next step should be to choose the best technique for analysing the model in RStudio. Due to fact that Pooled OLS (*pooling*) may have difficulties with panel data analysis, we also consider Fixed Effect (*within*) and Random effects (*random*) models. In order to determine which estimation technique will have the best performance for our data, we construct the models and, thereafter, run statistical tests in RStudio.

At first, we perform Chow test *pooltest* in RStudio, and thus test the null hypothesis that Pooled OLS is the best fit against the alternative hypothesis that Fixed Effects model should be chosen. The p-value produced by test is much less than the significance level, therefore we reject the null hypotheses and should exploit Fixed Effects model. Here and further on we will use value 0,05 as the significance level. Thereafter, we run Hausman *phptest* test, in order to identify if Fixed Effects or Random Effects model would be a better fit, with the null hypothesis that we should choose Fixed Effects. Expectedly, Random Effects model would match our analysis purposes better, because unlike Fixed Effects, it keeps all dummy variables estimates. The p-value produced by Hausman test is larger than significance level, therefore, Random model is obviously the preferred one. Table 9 summarizes results of both tests.

Table 9: Results of Chow and Hausman tests for the general models

	p-value	
	Models 07.09	Models 11.09
Chow test	0.000	0.000
Hausman test	0.8919	0.6576

However, to avoid misleading results and make the estimation process more precise, we should make sure that OLS assumptions of homoskedasticity and absence of serial correlation are kept.

In order to detect heteroskedasticity, we should run Breusch-Pagan *bptest* test in RStudio, with the null hypotheses that the residuals are homoscedastic. Due to extremely low p-value, it

becomes evident that the collected data is prone to heteroskedasticity. In order to investigate if, at the same time, serial correlation is present in the model, we conduct Durbin-Watson test *pwdtest* which tests the null hypothesis that observations are not correlated with an error term. As a result, we conclude that the model also violates assumption of serial correlation absence, because p-value is lower than the significance level (see table 10).

Table 10: Results of Breusch-Pagan and Durbin-Watson tests for the general models

	p-value	
	Model 07.09	Model 11.09
Breusch-Pagan test for heteroskedasticity	0.000	0.000
Durbin-Watson test for	0.000	0.000

The model has already been controlled for multicollinearity however, we are willing to obtain Variance Inflation Factor (VIF) values in RStudio in order to confirm our findings. VIF-value of 1 to 5 for any variable shows that there is no multicollinearity, while value of 5 or higher signifies the opposite. The results do not show that the modified model is prone to multicollinearity (see table 11).

Table 11: VIF values for the general models

Variable	VIF	
	Model 07.09	Model 11.09
<i>DURATION</i>	1.131409	1.162766
<i>TIME_TO_DEPARTURE</i>	1.000038	1.000196
<i>COMPETITION_DEPARTURE</i>	1.187649	1.221019
<i>MARKET_SHARE</i>	1.318942	1.307859
<i>early</i>	1.127334	1.126247
<i>late</i>	1.182075	1.241893
<i>early_evening</i>	1.239288	1.241395
<i>late_reg</i>	1.000002	1.000000
<i>lug_2</i>	2.261397	2.243928
<i>reb_2</i>	1.938031	1.919956
<i>ref_2</i>	3.071160	3.004132

Due to rejection of null hypotheses of homoskedasticity and no autocorrelation, we conclude that it would be more appropriate to estimate our data using the GLS method, so that assumption about absence of serial correlation, as well as homoskedasticity assumption are relaxed.

Table 12 summarizes the results of GLS estimation. It can be observed that majority of p-values are very close to zero. This can be the case due to fact that there are many identical observations in the collected data, and variability is low compared to the dataset size. In analysis sense it means that the corresponding coefficients are statistically significant.

The value of each statistically significant variable coefficient can be interpreted in the following way: by how much dependent variable changes with one-unit change in this explanatory variable, while holding all other explanatory variables constant. We will further on assume this when interpreting the regression coefficients. The sign of the coefficient value is important for our analysis, because it defines if there is positive or negative correlation between the specific independent variable and dependent variable, which is what we need to know to answer our research question. We will further on provide summary tables of the regression models, so that numbers obtained during the analysis are visible and results are transparent. We will focus not on explanation of Standard Errors or t-values, but on p-values.

For both models, “07.09” and “11.09”, the results of analysis show that coefficients of all variables are statistically significant, except for *late_reg*. p-value associated with *late_reg* is larger than the significance level 0,05 and thus insignificant. This means that our data does not provide a support for the hypothesis that time of the day one books tickets have an impact on the average price that is shown to the customer.

Our findings for variable *duration* suggest that there is a positive correlation between flight duration and the average ticket prices for both “07.09” and “11.09” models. According to our analysis of model “07.09”, while holding other independent variables constant, a one unit increase in duration raises the ticket price, on average, by 4,3837. It can therefore be said that one-minute increase in flight length increases the ticket price, on average, by 4,4 NOK. Model “11.09”, in turn, shows a 3,1 NOK increase.

In both models, *time_to_departure* appears to be negative correlated with ticket price, that is consistent with economic literature. In other words, our data support that the fares, on average,

increase by 9,2 NOK per day prior to departure day according to “07.09” dataset, and by 13,3 NOK according to “11.09” dataset.

market_share, in turn, appears to be positive correlated with average price. It should be stated that market share values are expressed in decimal numbers, so the coefficients values for market share should be correspondingly adjusted in case one willing to interpret the value precisely.

Coefficient values of dummy variables represent the difference among the experimental group and comparison group in terms of dependent variable. For both models, the coefficient for *competition_departure* is negative, so the data provides an evidence that in case there is a flight provided by competitor in a time range of 30 minutes before or after the flight departure, ticket prices are, on average, lower.

Coefficients for ticket restriction variables *lug_2*, *reb_2* and *ref_2* are consistent with logical sense and increase ticket price according to the data. The largest difference between reference and comparison group is observed for variable *reb_2*, which can be explained by fact only the most expensive fares of both airlines include rebooking option.

So far, both “07.09” and “11.09” models have represented the similar trends in terms of coefficient signs. However, this is not the case for variables *early*, *late* and *late_evening*. Considering “07.09” dataset, we find support that tickets are on average more expensive for early flights (81 NOK more) compared to the rest of the day, as well as for late flights (128 NOK more) compared to the rest of the day, while all flights departing after 16:00 are on average cheaper than flights departing earlier. The situation is different for “11.09” dataset. This dataset does not provide an evidence that late flight tickets are on average more expensive or cheap than the rest of the day. However, we find the support for that early flight tickets are on average cheaper than the rest of the day (28 NOK less) and that the tickets for flights that depart after 16:00 are on average more expensive (35,6 NOK more). Different results could be observed due to weekday effect (departure on different weekdays).

Table 12: Summary of the general regression results

Variable	Model	Value	Std.Error	t-value	p-value
(Intercept)	07.09	-230.6888	15.395327	-14.98434	0.0000
	11.09	-35.6181	15.394425	-2.31370	0.0207

<i>DURATION</i>	07.09	4.3837	0.151733	28.89123	0.0000
	11.09	3.1290	0.149766	20.89247	0.0000
<i>TIME_TO_DEPARTURE</i>	07.09	-9.2440	0.325029	-28.44056	0.0000
	11.09	-13.2912	0.328754	-40.42897	0.0000
<i>COMPETITION_DEPARTURE</i>	07.09	-85.8257	6.108940	-14.04920	0.0000
	11.09	-68.2145	6.192263	-11.01608	0.0000
<i>MARKET_SHARE</i>	07.09	1945.9763	23.028629	84.50248	0.0000
	11.09	1979.0166	22.274301	88.84753	0.0000
<i>early</i>	07.09	81.0016	8.703413	9.30688	0.0015
	11.09	-28.0307	8.830275	-3.17439	0.0000
<i>late</i>	07.09	127.9503	13.453789	9.51036	0.0000
	11.09	-19.5714	15.206760	-1.28702	0.1981
<i>early_evening</i>	07.09	-15.3657	6.445343	-2.38400	0.0171
	11.09	35.5576	6.491460	5.47760	0.0000
<i>late_reg</i>	07.09	-1.1048	5.589082	-0.19767	0.1860
	11.09	7.4772	5.653241	1.32264	0.0636
<i>lug_2</i>	07.09	388.8783	9.359720	41.54807	0.0000
	11.09	285.1387	10.110227	28.20299	0.0000
<i>lug_2</i>	07.09	751.0274	8.469849	88.67070	0.0000
	11.09	779.5868	8.463861	92.10770	0.0000
<i>ref_2</i>	07.09	398.6491	10.497001	37.97742	0.0000
	11.09	425.1757	11.090932	38.33544	0.0000

4.3 The “Economy subset” analysis

In this thesis we also propose model which only considers data collected for specific fare categories provided by both Norwegian and SAS. In order to analyse our dataset in such way we create a data subset of the following ticket categories: Norwegian LowFare+ and SAS Go Smart observations only. We are therefore interested how the results of such analysis would be different to the ones observed in general model.

The logic behind choosing these ticket categories for the model is that they are well-comparable and could be substitutes for one another. Both fares can be attributed to economy class and are equal in such characteristics as one unit of luggage included in price and excluded refundability/rebooking options. There is, for instance, no alternative to Norwegian LowFare, the

fare which assumes hand-luggage only, provided by SAS. On the other hand, fares SAS Plus Smart and SAS Plus Pro cannot be compared to any ticket provided by Norwegian in the best way. The model which compares Norwegian Flex and SAS Plus Full Flex could also be analysed, because those travel classes are well-substitutes to each other in terms of free refundability and rebooking options. However, it was observed that prices for SAS Plus Full Flex fare did not change depending on hours of flight departure and time to departure, and it is most probably the case due to free rebooking possibility included in price. Norwegian Flex ticket prices, in turn, followed the pattern to change proportionally to other fares provided by Norwegian, however, the low variability was observed. This would make it purposeless to estimate such a model in terms of such variables as *competition_departure*, *time_to_departure*, *early*, *late* and *early_evening*.

The following model is proposed to analyze the data subset:

$$\begin{aligned}
 \text{ticket_price} = & \beta_0 + \beta_1 \text{duration} + \beta_2 \text{time_to_departure} + \beta_3 \text{market_share} + \beta_4 \\
 & \text{competition_departure} + \beta_5 \text{early} + \beta_6 \text{late} + \beta_7 \text{early_evening} + \beta_8 \text{late_reg}
 \end{aligned}$$

(4.3.1)

We apply the identical model to both 07.09 and 11.09 datasets and use the same test algorithm as for the general model. As a result, GLS estimation still appears to be preferable.

Table 13 shows the summary of the estimation. It can be observed that all coefficients except for *late_reg* are statistically significant, for both models.

The “economy” data subsets of both “07.09” and “11.09” datasets, support that economy-ticket prices on average increase with flight duration. There is also an evidence that the economy fares become more expensive as date of departure gets closer, due to the coefficients sign for *time_of_departure*. If we look at the market-share coefficient value, we see that there is an evidence of positive correlation between market share and price of the economy-class ticket.

As for the general model, we find the support for early flights to be on average more expensive than rest of the day on 07.09, and on average cheaper than rest of the day on 11.09. Coefficient value for *late* stays positive for “07.09” model, so we find the evidence that economy tickets for late flights are on average more expensive on that day. It can also be noticed that coefficient for *late* for “11.09” model, which was not significant in the general “11.09” model, now appears to

be significant and negative. We can therefore say that our data supports that economy-ticket prices are on average cheaper for the evening flights on 11th September 2020.

Regression coefficient value for *early_evening* is positive for both days. So, we conclude that our datasets support that economy tickets are on average more expensive for flights that depart after 16:00 compared to the earlier flights both on 7th and 11th September. Moreover, the difference observed for 11.09 is bigger (53 versus 121,5).

Table 13: Summary of the “Economy subset” regression results

Variables	Model	Value	Std.Error	t-value	p-value
(Intercept)	07.09	150.0371	15.530925	9.66054	0.0000
	11.09	472.3143	17.240951	27.39491	0.0000
<i>DURATION</i>	07.09	6.7411	0.151068	44.62280	0.0000
	11.09	5.2975	0.166658	31.78665	0.0000
<i>TIME_TO_DEPARTURE</i>	07.09	-14.7900	0.332323	-44.50486	0.0000
	11.09	-19.2574	0.370556	-51.96894	0.0000
<i>COMPETITION_DEPARTURE</i>	07.09	-74.4000	6.295770	-11.81746	0.0000
	11.09	-51.4513	7.068910	-7.27854	0.0000
<i>MARKET_SHARE</i>	07.09	1071.203	22.452618	47.70950	0.0000
	11.09	871.6926	23.986818	36.34048	0.0000
<i>early</i>	07.09	181.8963	8.921789	20.38787	0.0000
	11.09	-28.3191	-28.3191	-2.78949	0.0053
<i>late</i>	07.09	34.6894	13.808403	2.51220	0.0120
	11.09	-188.9333	17.203900	-10.98200	0.0000
<i>early_evening</i>	07.09	52.9766	6.606290	8.01911	0.0000
	11.09	121.4839	7.331000	16.57125	0.0000
<i>late_reg</i>	07.09	-3.3214	5.744952	-0.57814	0.5632
	11.09	10.0999	6.426028	1.57172	0.1161

4.4 “Norwegian” and “SAS” analysis

According to economic literature, low-cost airlines and network airlines may use different strategies in terms of, for example, their market share on a route and degree of competition.

Considering fact that Norwegian is a low-cost carrier while SAS is not, it is reasonable to also analyze datasets with Norwegian-data and SAS-data separately. We considered the option to generate a dummy variable that distinguishes among two air companies, and add interaction terms to the general model, however, GLS, which was chosen as the estimation method for our data, appeared to have difficulties while analyzing model of this type. Therefore, it was decided to proceed with analysis of Norwegian and SAS datasets independently. We applied the procedure with statistical tests for both models and, as a result, favored GLS estimation. Again, we created the identical models for analysis of data for both departure days, 7th September 2020 and 11th September 2020.

The regression equation for “Norwegian” models:

$$\begin{aligned}
 \text{ticket_price} = & \beta_0 + \beta_1 \text{duration} + \beta_2 \text{time_to_departure} + \beta_3 \text{market_share} + \beta_4 \\
 & \text{competition_departure} + \beta_5 \text{early} + \beta_6 \text{late} + \beta_7 \text{early_evening} + \beta_8 \text{late_reg} + \beta_9 \text{ref_2} \\
 & (4.4.1)
 \end{aligned}$$

The model would have a multicollinearity problem, if all three ticket-restrictions dummy variables *lug_2*, *reb_2* and *ref_2* were kept. The reason is that these variables take value 1 for Norwegian Flex fare only, while other ticket categories provided by Norwegian are economy-class tickets which do not include two units of luggage, rebooking and refundability option. It means that we can only keep one of the three variables, because any of those will show the same effect. We, therefore, only leave *ref_2* in our model.

The regression equation for “SAS” model:

$$\begin{aligned}
 \text{ticket_price} = & \beta_0 + \beta_1 \text{duration} + \beta_2 \text{time_to_departure} + \beta_3 \text{market_share} + \beta_4 \\
 & \text{competition_departure} + \beta_5 \text{early} + \beta_6 \text{late} + \beta_7 \text{early_evening} + \beta_8 \text{late_reg} + \beta_9 \text{lug_2} + \\
 & \beta_{10} \text{reb_2} + \beta_{11} \text{ref_2} \quad (4.4.2)
 \end{aligned}$$

Unlike Norwegian fares, SAS ticket categories cannot be distinguished by only one variable out of *lug_2*, *reb_2* and *ref_2*. SAS Plus Smart, SAS Plus Pro and SAS Plus Full Flex are different in terms of those variables, therefore, every variable can be kept in the model, in order that the variation in average price is better explained.

Table 14 summarizes the regression results for Norwegian models and table 15 for the SAS models. We find the evidence that flight duration and average ticket price for both SAS and Norwegian are positively correlated. The results show there is on average a 4,2 NOK increase in ticket price on Norwegian with one-minute increase in flight length, according to “07.09” model, and a 4,3 NOK increase according to “11.09” model. There is, in turn, 6,3 NOK increase in SAS ticket price for one-minute increase in flight duration according to “07.09” model, and 5,5 NOK increase according to “11.09” model.

Coefficient sign for *time_to_departure* is consistent with both literature and our previous findings. We thus find a support for both Norwegian and SAS average ticket price to increase as departure date is closer.

However, Norwegian and SAS models show different results in terms of coefficient value for *competition_departure*. Both “07.09” and “11.09” models for Norwegian show negative correlation between *competition_departure* and average ticket price. Model “07.09” for SAS suggests positive correlation between those two variables, but dataset for 11th September for SAS does not provide enough evidence that those variables are correlated (p-value is larger than 0,05 for *competition_departure* in “11.09” model).

Considering variable *market_share*, both SAS datasets support that increased market share leads to higher price of SAS ticket on average. The coefficient for *market_share* for Norwegian model for 7th September is negative, and it means that decrease in average Norwegian ticket price with increased market share is observed. Norwegian dataset for 11th September, in turn, provides the evidence, that on average, price becomes higher with increased market share, and that is consistent with both results for SAS datasets and our previous findings.

According to the signs of coefficient value for *early*, both Norwegian and SAS models for 7th September give the evidence that early flight tickets for this day, on average, are more expensive compared to the rest of the day. Both models for 11th September, in turn, show the negative correlation, so our datasets support that unlike 7th September, early flight tickets appear to be cheaper compared to rest of the 11th September. This finding is consistent with previous results for general and “economy subset” models.

We do not find an evidence that variable *late* is anyhow correlated with average ticket price on Norwegian for 7th September. However, the coefficient value for *late* in Norwegian model for 11.09 is significant and negative, so there is a support that the tickets were cheaper compared to rest of the day on 11th September. Considering SAS models, we find the evidence that late flight tickets are more expensive compared to rest of the day on 7th September, and cheaper compared to the rest of the day on 11th September.

Both Norwegian and SAS models for both dates lead to result that variable *early_evening* is positive correlated with average ticket price. It means that we find the support for average ticket price to be more expensive for flights that depart after 16:00 compared to the earlier flights, regardless of which day and which airline company of our sample is considered.

It is interesting that coefficient for variable *late_reg* is significant and positive in model for Norwegian for 11th September. This dataset for Norwegian, thus, provide the evidence that Norwegian ticket prices for flights with departure on 11th September, registered during late evening were 14,8 NOK more expensive compared to the prices registered during daytime. However, in all other models, *late_reg* coefficient value is insignificant.

Expectedly, coefficients for *lug_2*, *reb_2* and *ref_2* are significant and positive for SAS models, as it expected to be. *ref_2* coefficient values for Norwegian models are significant, and shows that, on average, refundability option raises the ticket price by 555,4 NOK according to 07.09 model and by 510,1 NOK according to 11.09 model. Due to fact that Norwegian Flex is the only one out of Norwegian ticket categories that includes two units of luggage, as well as rebooking and refundability option, the effect would be the same for *lug_2* and *reb_2*.

Table 14: Summary of the “Norwegian” regression results

Variables	Model	Value	Std.Error	t-value	p-value
(Intercept)	07.09	582.5215	9.472958	61.49310	0.0000
	11.09	653.0426	15.936557	40.97764	0.0000
<i>DURATION</i>	07.09	4.2227	0.082504	51.18181	0.0000
	11.09	4.3477	0.134967	32.2129	0.0000
<i>TIME_TO_DEPARTURE</i>	07.09	-9.5587	0.166296	-57.47989	0.0000
	11.09	-16.9141	0.281958	-59.98794	0.0000
<i>COMPETITION_DEPARTURE</i>	07.09	-13.0297	3.489228	-3.73426	0.0002
	11.09	-46.6448	5.940731	-7.85170	0.0000
<i>MARKET_SHARE</i>	07.09	-43.8542	14.644398	-2.99461	0.0028
	11.09	349.5709	24.971539	13.99877	0.0000
<i>early</i>	07.09	123.6062	4.517043	27.36441	0.0000
	11.09	-34.2997	7.984594	-4.29573	0.0000
<i>late</i>	07.09	-1.9811	8.170410	-0.24248	0.8084
	11.09	-185.4304	15.028685	-12.33843	0.0000
<i>early_evening</i>	07.09	26.9663	3.253598	8.28813	0.0000
	11.09	103.2316	5.492738	18.79419	0.0000
<i>late_reg</i>	07.09	-0.3313	2.872637	-0.11533	0.9082
	11.09	14.8276	4.884603	3.03557	0.0024
<i>ref_2</i>	07.09	555.3718	3.046860	182.27680	0.0000
	11.09	510.1466	5.176565	98.54925	0.0000

Table 15: Summary of the “SAS” regression results

Variables	Model	Value	Std.Error	t-value	p-value
(Intercept)	07.09	77.2888	10.625847	7.27366	0.0000
	11.09	263.9500	11.287875	23.38349	0.0000
<i>DURATION</i>	07.09	6.3029	0.104060	60.57018	0.0000
	11.09	5.4509	0.109219	49.90751	0.0000
<i>TIME_TO_DEPARTURE</i>	07.09	-8.9317	0.207866	-42.96878	0.0000
	11.09	-11.2680	0.220184	-51.17530	0.0000
<i>COMPETITION_DEPARTURE</i>	07.09	11.2578	3.898153	2.88799	0.0039
	11.09	6.4210	4.152500	1.54631	0.1220
<i>MARKET_SHARE</i>	07.09	1086.9349	17.587922	61.80007	0.0000
	11.09	992.2173	18.565218	53.44496	0.0000
<i>early</i>	07.09	33.3555	5.658701	5.89455	0.0000
	11.09	-43.6595	5.891593	-7.41047	0.0000
<i>late</i>	07.09	48.3630	8.196531	5.90042	0.0000
	11.09	-71.6938	9.731566	-7.36714	0.0000
<i>early_evening</i>	07.09	22.6732	4.193005	5.40739	0.0000
	11.09	55.6723	4.416654	12.60509	0.0000
<i>late_reg</i>	07.09	-1.5545	3.565212	-0.43603	0.6628
	11.09	3.4814	3.769712	0.92351	0.3558
<i>lug_2</i>	07.09	394.0432	5.439089	72.44654	0.0000
	11.09	344.6770	6.017788	57.27637	0.0000
<i>reb_2</i>	07.09	1362.8543	4.822284	282.61595	0.0000
	11.09	1342.5287	5.020419	267.41367	0.0000
<i>ref_2</i>	07.09	419.4759	5.438323	77.13332	0.0000
	11.09	443.3832	6.017788	73.67877	0.0000

5. Concluding discussion

The analysis conducted in this thesis used linear regression and GLS method to test the multiple null hypotheses that flight duration, time to departure, competition with regard to departure, market share of the airline on route, early departure (before 08:00AM), late departure (after 20:00PM), departure after 16:00PM, late registration, as well as basic ticket characteristics are not correlated with the fare price against the alternative hypotheses that there is a correlation.

For each of our samples, the one for flights with departure on 7th September 2020 and the one for flights with departure on 11th September 2020, four models were examined. The general model analysed all fares and both Norwegian and SAS, the “subset” model – only Norwegian LowFare+ and SAS Go Smart economy fares, the “Norwegian” - purely fares of Norwegian Air Shuttle and “SAS” model – only SAS fares.

Regardless which model is applied, our data provide enough evidence that flight duration is correlated with ticket price, and this correlation is positive. According to the results, one unit increase in flight length (one-minute increase) increases average ticket price by value from 3,1 to 6,7 NOK depending on the model.

In terms of correlation between time to departure and fare price, our research is consistent with economic literature such as Pels and Rietveld, 2004, Gaggero and Piga, 2010 and others. Our data gives support to the correlation between these two variables. Regardless of the model, there is an evidence that prices on average increase over time till date of departure. The range of increase in average price is from 8,9 to 19,3 NOK per day, depending on the model. We therefore conclude that, according to our dataset, inter-temporal price discrimination takes place on airline market in Norway despite the complicated situation caused by COVID-pandemic.

In general, it was observed that the larger is market share of the airline on the route, the higher are airline’s average fare prices. However, the “Norwegian” model which analysed the sample for 7th September suggests that average ticket price decreases with increased market share. The results for Norwegian Air Shuttle samples are, therefore, ambiguous.

According to both general and “economy subset” models regardless of the sample used, if there is a competitor’s flight available in time range of 30 minutes before or after the departure, the average ticket price tends to decrease. However, when we apply “Norwegian” and “SAS”

models and Norwegian- and SAS-datasets are analysed separately, it can be seemed that, unlike Norwegian, SAS prices, vice versa, tend to increase when there is such a competition with regards to departure. This may be the case due to differences among low-cost and network strategies.

Regression coefficients for dummy variables which define the fare characteristics, such as two units of luggage, rebooking possibility and refundability option, are consistent with realistic expectations regardless of the model they were included in. Free rebooking option is the service for which client of SAS needs to pay the most. For customer who chooses Norwegian, two units of luggage, rebooking and refundability would be the services which are equally costly.

Another aspect we were interested in was flight characteristics in terms of its time of departure. Our data provides enough evidence that early flights with departure before 8AM were on average more expensive than rest of the day on Monday 7th September in general, for economy ticket subset and for each airline company. On the other hand, there is an evidence that such flights were on average cheaper than the rest of the day on Friday 11th September, regardless of which model we apply. The same pattern is observed for late flights with departure after 8PM, and the only exception is “Norwegian” model for 7th September. Our dataset did not provide enough evidence that late flights on 7th September were more expensive or cheap than rest of that day. According to our data, there is a possibility that “weekday” effect takes place.

If we apply “Norwegian”, “SAS” or “economy subset” models to our datasets, we find the evidence that flights with departure after 16:00 were on average more expensive than rest of the day both on 7th and 11th of September. However, the numeric change is larger for Friday, 11th September. Moreover, the general model does not show the same pattern for 7th of September: flights with departure after 16:00 appear to be cheaper compared to the rest of the day. This may occur due to weekday effect as well. Tickets on Friday may be relatively cheap in the early morning and late evening compared to rest of the day, however, relatively costly for flights after 16:00 compared to earlier flights. There is a possibility that the pattern is different for Monday.

In general, we do not find support for time of the day when the prices are registered to be correlated with ticket price which is shown on the website. Only in “Norwegian” model for 7th September the coefficient for dummy variable which distinguishes among day and evening

registration appears to be statistically significant. Thinking critically, the guess on possibility such correlation may not be well-grounded.

Summarized, the core assumptions, such as increase of prices over time to departure date, positive correlation between flight duration and additional service with fare price, find support in our datasets. However, some other results are ambiguous. The assumptions and conclusions we made during the analysis could take places according to our results, however, we believe that extended data samples are required to conduct more accurate research. For instance, we cannot be sure if the weekday effect occurs due to different customer behaviour on Monday and Friday, because it could be different on Monday and another weekday as well. Nevertheless, our data supports that pricing strategies of airlines in Norway are somehow different for two different days of the same week. The fact that domestic airline market in Norway is currently not very stable due to COVID-19 pandemic, could have impact on results of the analysis. Further research could potentially include extended data sample and be performed during more stable period, as well as account for changes on Norwegian airline market, such as appearance of Wizz Air airlines etc. Such research could also be compared to our analysis in terms of findings.

6. List of references

- Alderighi, M., et al. (2011). "A case study of pricing strategies in European airline markets: The London–Amsterdam route." Journal of Air Transport Management **17**(6): 369-373.
- Alexopoulos, E. C. (2010). "Introduction to multivariate regression analysis." Hippokratia **14**(Suppl 1): 23.
- Avinor.no (2020). "Traffic statistics on passengers and flight movements for the Avinor Group." from <https://avinor.no/en/corporate/about-us/statistics/traffic-statistics>.
- Baltagi, B. H. (2011). Generalized Least Squares. Econometrics, Springer: 223-239.
- Bilotkach, V., et al. (2010). "Are airlines' price-setting strategies different?" Journal of Air Transport Management **16**(1): 1-6.
- Borenstein, S. and N. L. Rose (1994). "Competition and price dispersion in the US airline industry." Journal of Political Economy **102**(4): 653-683.
- Brekke, K., et al. (2019). "Samfunnsøkonomen: Effekter av konkurransereglene for norske forbrukere." from <https://konkurransetilsynet.no/effekter-av-konkurransereglene-for-norske-forbrukere/>.
- Breusch, T. S. and A. R. Pagan (1979). "A simple test for heteroscedasticity and random coefficient variation." Econometrica: Journal of the Econometric Society: 1287-1294.
- Chakrabarty, D. and L. Kutlu (2014). "Competition and price dispersion in the airline markets." Applied Economics **46**(28): 3421-3436.
- Chen, M. and Z. L. Chen (2015). "Recent developments in dynamic pricing research: multiple products, competition, and limited demand information." Production and Operations Management **24**(5): 704-731.
- Clark, R. and N. Vincent (2012). "Capacity-contingent pricing and competition in the airline industry." Journal of Air Transport Management **24**: 7-11.
- Croissant, Y. and G. Millo (2008). "Panel data econometrics in R: The plm package." Journal of statistical software **27**(2): 1-43.

Diagilev, E. (2020). "10 Largest Airlines in Europe by Passenger Numbers." 2020, from https://www.skycop.com/fun-facts/10-largest-airlines-in-europe-by-passenger-numbers/?utm_source=facebook&utm_medium=brand&utm_campaign=DEF-1119&utm_content=post.

Escobari, D. and P. Jindapon (2014). "Price discrimination through refund contracts in airlines." International Journal of Industrial Organization **34**: 1-8.

Faruqui, A. (2012). The ethics of dynamic pricing. Smart Grid, Elsevier: 61-83.

Ferreira, J. C. and C. M. Patino (2015). "What does the p value really mean?" Jornal Brasileiro de Pneumologia.

Franke, M. (2004). "Competition between network carriers and low-cost carriers—retreat battle or breakthrough to a new level of efficiency?" Journal of Air Transport Management **10**(1): 15-21.

Frost, J. 7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression.

Gaggero, A. A. and C. A. Piga (2010). "Airline competition in the British Isles." Transportation Research Part E: Logistics and Transportation Review **46**(2): 270-279.

Gao, J., et al. (2020). "Heterogeneous panel data models with cross-sectional dependence." Journal of Econometrics.

Giaume, S. and S. Guillou (2004). "Price discrimination and concentration in European airline markets." Journal of Air Transport Management **10**(5): 305-310.

Halvorsen, B. S. and M. R. Skaug (2016). Bonusprogrammer i luftfartsmarkedet: en empirisk analyse av SAS Eurobonus.

Haouaria, A. (2020). "From Full Service to Low Cost and Charters: How Different Categories of Airlines work ?". from <http://www.aviation-stars.com/crew-topics/crew-topics/article/from-full-service%2%A0to-low-cost-and-charters-how-different-categories-of-airlines-work-by-ahmed-haouaria-19.html>.

Konkurransetilsynet (2020). "Midlertidig unntak fra konkurranseloven § 10." from <https://konkurransetilsynet.no/midlertidig-unntak-fra-konkurranseloven-%C2%A710/>.

Malighetti, P., et al. (2009). "Pricing strategies of low-cost airlines: The Ryanair case study." Journal of Air Transport Management **15**(4): 195-203.

Mazareanu, E. (2020). "Low-cost carrier market share in Europe 2009-2019." from <https://www.statista.com/statistics/1117218/low-cost-carrier-market-share-europe/>.

Muthama Musau, V., et al. (2015). "Modeling Panel Data: Comparison of GLS Estimation and Robust Covariance Matrix Estimation." American Journal of Theoretical and Applied Statistics **Volume 4**(Issue 3): 185-191.

norwegian.com. "About us." 2020, from <https://www.norwegian.com/uk/about/>.

norwegian.com (2020). "Norwegian Reward." from <https://en.norwegianreward.com/about-us/how-it-works>.

norwegian.com (2020). "Our ticket types." from <https://www.norwegian.com/us/booking/booking-information/ticket-types/>.

norwegian.no. "År for år." 2020, from <https://www.norwegian.no/om-oss/var-historie/ar-for-ar/>.

norwegian.no. "Om Oss. Hva er Norwegian Reward." 2020, from https://no.norwegianreward.com/om-oss?utm_source=norwegian&utm_medium=web&utm_campaign=nas-footer.

norwegian.no. "Priser og utmerkelse." 2020, from <https://www.norwegian.no/om-oss/var-historie/priser-og-utmerkelse/>.

norwegian.no. "Vår visjon og verdier." 2020, from <https://www.norwegian.no/om-oss/var-historie/visjon-og-verdier/>.

Park, H. M. (2011). "Practical Guides To Panel Data Modeling: A Step by Step Analysis Using Stata." from https://www.academia.edu/31700251/Practical_Guides_To_Panel_Data_Modeling_A_Step_by_Step.

Pels, E. and P. Rietveld (2004). "Airline pricing behaviour in the London–Paris market." Journal of Air Transport Management **10**(4): 277-281.

Pettinger, T. (2017). "Airline price discrimination." from <https://www.economicshelp.org/blog/7767/business/airline-price-discrimination/>.

Regjeringen.no (2019). "Det norske luftfartsmarkedet." from <https://www.regjeringen.no/no/dokumenter/nou-2019-22/id2680751/?ch=5>.

Robert J Casson DPhil FRANZCO and L. D. F. MBBS1 (2014). "Understanding and checking the assumptions of

linear regression: a primer for medical researchers." from <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ceo.12358>.

sas.com. "About SAS." 2020, from <https://www.sasgroup.net/>.

sas.com. "History milestones." 2020, from <https://www.sasgroup.net/about-sas/sas-history/history-milestones/>.

sas.com (2020). "SAS EuroBonus." from <https://www.flysas.com/en/eurobonus/>.

sas.com (2020). "Ticket types." from <https://www.flysas.com/en/fly-with-us/ticket-options/ticket-types/>.

SSB.no (2019). "Lufttransport statistikk."

Strauss, A. K., et al. (2018). "A review of choice-based revenue management: Theory and methods." European Journal of Operational Research **271**(2): 375-387.

Taboga, M. (2017). "Generalized least squares." Lectures on probability theory and mathematical statistics, Third edition. Kindle Direct Publishing. Online appendix. from <https://www.statlect.com/fundamentals-of-statistics/generalized-least-squares>.

Varian, H. R. (1989). "Price discrimination." Handbook of industrial organization **1**: 597-654.

Vege, H. H. (2020). "Norwegian svarer i flykrigen - dumper billettprisene til bunns." from <https://www.nettavisen.no/okonomi/norwegian-svarer-i-flykrigen-dumper-billettprisene-til-bunns/s/12-95-3424028699>.

Wooldridge, J. M. (2019). "Correlated random effects models with unbalanced panels." Journal of Econometrics **211**(1): 137-150.

Yobero, C. (2016). "Methods for Detecting and Resolving Heteroskedasticity." from https://rstudio-pubs-static.s3.amazonaws.com/187387_3ca34c107405427db0e0f01252b3fbdb.html#:~:text=One%20informa%20way%20of%20detecting,plot%2C%20then%20heteroskedasticity%20is%20present.

(Frost , norwegian.com , norwegian.no , sas.com , Breusch and Pagan 1979, Borenstein and Rose 1994, Franke 2004, Croissant and Millo 2008, Alexopoulos 2010, Bilotkach, Gorodnichenko et al. 2010, Baltagi 2011, Park 2011, Chakrabarty and Kutlu 2014, Robert J Casson DPhil FRANZCO and MBBS1 2014, Chen and Chen 2015, Ferreira and Patino 2015, Muthama Musau, Gichuhi Waititu et al. 2015, Yobero 2016, Pettinger 2017, Taboga 2017, Brekke, Sjørgard et al. 2019, Gao, Xia et al. 2020, Haouaria 2020, norwegian.com 2020, sas.com 2020)

7. Appendix

Market shares of the airlines

Table 16: Market shares of the airlines on domestic routes in Norway on 07.09.2020

Route	Market shares 07.09.2020			
	SAS	Norwegian	Widerøe	Total
Oslo - Stavanger	0.529	0.471	-	1.000
Stavanger - Oslo	0.529	0.471	-	1.000
Oslo - Trondheim	0.529	0.471	-	1.000
Trondheim - Oslo	0.529	0.471	-	1.000
Oslo - Bergen	0.4737	0.4737	0.0526	1.000
Bergen - Oslo	0.4737	0.4737	0.0526	1.000
Oslo - Tromsø	0.666	0.333	-	1.000
Tromsø - Oslo	0.666	0.333	-	1.000
Stavanger - Trondheim	1.000	-	-	1.000
Trondheim - Stavanger	1.000	-	-	1.000
Stavanger - Bergen	0.500	0.143	0.357	1.000
Bergen - Stavanger	0.500	0.143	0.357	1.000
Trondheim - Bergen	0.714	0.143	0.143	1.000
Bergen - Trondheim	0.714	0.143	0.143	1.000

Table 17: Market shares of the airlines on domestic routes in Norway on 11.09.2020

Route	Market shares 11.09.2020			
	SAS	Norwegian	Widerøe	Total
Oslo - Stavanger	0.588	0.412	-	1.000
Stavanger - Oslo	0.529	0.471	-	1.000
Oslo - Trondheim	0.5	0.5	-	1.000
Trondheim - Oslo	0.529	0.471	-	1.000
Oslo - Bergen	0.526	0.421	0.053	1.000
Bergen - Oslo	0.4737	0.4737	0.0526	1.000
Oslo - Tromsø	0.7	0.3	-	1.000
Tromsø - Oslo	0.666	0.333	-	1.000
Stavanger - Trondheim	1.000	-	-	1.000
Trondheim - Stavanger	1.000	-	-	1.000

Stavanger - Bergen	0.500	0.143	0.357	1.000
Bergen - Stavanger	0.500	0.143	0.357	1.000
Trondheim - Bergen	0.714	0.143	0.143	1.000
Bergen - Trondheim	0.714	0.143	0.143	1.000

The route samples

Norwegian Air Shuttle (Norwegian) sample

FLIGHT	DEPARTURE
OSLO_STAVANGER	9/7/2020 8:00
OSLO_STAVANGER	9/7/2020 9:00
OSLO_STAVANGER	9/7/2020 13:30
OSLO_STAVANGER	9/7/2020 15:00
OSLO_STAVANGER	9/7/2020 16:20
OSLO_STAVANGER	9/7/2020 17:15
OSLO_STAVANGER	9/7/2020 19:55
OSLO_STAVANGER	9/7/2020 20:40
STAVANGER_OSLO	9/7/2020 7:30
STAVANGER_OSLO	9/7/2020 8:30
STAVANGER_OSLO	9/7/2020 9:15
STAVANGER_OSLO	9/7/2020 10:20
STAVANGER_OSLO	9/7/2020 15:00
STAVANGER_OSLO	9/7/2020 16:30
STAVANGER_OSLO	9/7/2020 17:40
STAVANGER_OSLO	9/7/2020 18:35
OSLO_TRONDHEIM	9/7/2020 8:00
OSLO_TRONDHEIM	9/7/2020 8:55
OSLO_TRONDHEIM	9/7/2020 10:40
OSLO_TRONDHEIM	9/7/2020 13:20
OSLO_TRONDHEIM	9/7/2020 16:20
OSLO_TRONDHEIM	9/7/2020 17:40
OSLO_TRONDHEIM	9/7/2020 19:00
OSLO_TRONDHEIM	9/7/2020 20:20
TRONDHEIM_OSLO	9/7/2020 6:35
TRONDHEIM_OSLO	9/7/2020 7:30
TRONDHEIM_OSLO	9/7/2020 9:20
TRONDHEIM_OSLO	9/7/2020 12:00
TRONDHEIM_OSLO	9/7/2020 15:00
TRONDHEIM_OSLO	9/7/2020 16:20

FLIGHT	DEPARTURE
OSLO_STAVANGER	9/11/2020 8:00
OSLO_STAVANGER	9/11/2020 9:00
OSLO_STAVANGER	9/11/2020 13:30
OSLO_STAVANGER	9/11/2020 15:00
OSLO_STAVANGER	9/11/2020 16:20
OSLO_STAVANGER	9/11/2020 17:15
OSLO_STAVANGER	9/11/2020 19:55
STAVANGER_OSLO	9/11/2020 7:30
STAVANGER_OSLO	9/11/2020 8:30
STAVANGER_OSLO	9/11/2020 9:15
STAVANGER_OSLO	9/11/2020 10:20
STAVANGER_OSLO	9/11/2020 15:00
STAVANGER_OSLO	9/11/2020 16:30
STAVANGER_OSLO	9/11/2020 17:40
STAVANGER_OSLO	9/11/2020 18:35
OSLO_TRONDHEIM	9/11/2020 8:00
OSLO_TRONDHEIM	9/11/2020 8:55
OSLO_TRONDHEIM	9/11/2020 10:40
OSLO_TRONDHEIM	9/11/2020 13:20
OSLO_TRONDHEIM	9/11/2020 16:20
OSLO_TRONDHEIM	9/11/2020 17:40
OSLO_TRONDHEIM	9/11/2020 19:00
OSLO_TRONDHEIM	9/11/2020 20:20
TRONDHEIM_OSLO	9/11/2020 6:35
TRONDHEIM_OSLO	9/11/2020 7:30
TRONDHEIM_OSLO	9/11/2020 9:20
TRONDHEIM_OSLO	9/11/2020 12:00
TRONDHEIM_OSLO	9/11/2020 15:00
TRONDHEIM_OSLO	9/11/2020 16:20
TRONDHEIM_OSLO	9/11/2020 17:40

TRONDHEIM_OSLO	9/7/2020 17:40
TRONDHEIM_OSLO	9/7/2020 19:00
OSLO_BERGEN	9/7/2020 7:00
OSLO_BERGEN	9/7/2020 8:00
OSLO_BERGEN	9/7/2020 8:55
OSLO_BERGEN	9/7/2020 10:40
OSLO_BERGEN	9/7/2020 13:20
OSLO_BERGEN	9/7/2020 14:55
OSLO_BERGEN	9/7/2020 16:20
OSLO_BERGEN	9/7/2020 17:00
OSLO_BERGEN	9/7/2020 19:00
BERGEN_OSLO	9/7/2020 6:35
BERGEN_OSLO	9/7/2020 7:30
BERGEN_OSLO	9/7/2020 8:30
BERGEN_OSLO	9/7/2020 9:20
BERGEN_OSLO	9/7/2020 12:00
BERGEN_OSLO	9/7/2020 15:00
BERGEN_OSLO	9/7/2020 16:20
BERGEN_OSLO	9/7/2020 17:40
BERGEN_OSLO	9/7/2020 18:20
OSLO_TROMSO	9/7/2020 8:30
OSLO_TROMSO	9/7/2020 13:45
OSLO_TROMSO	9/7/2020 19:45
TROMSO_OSLO	9/7/2020 6:30
TROMSO_OSLO	9/7/2020 10:55
TROMSO_OSLO	9/7/2020 21:10
STAVANGER_BERGEN	9/7/2020 6:30
STAVANGER_BERGEN	9/7/2020 17:45
BERGEN_STAVANGER	9/7/2020 7:30
BERGEN_STAVANGER	9/7/2020 16:40
TRONDHEIM_BERGEN	9/7/2020 10:20
BERGEN_TRONDHEIM	9/7/2020 11:45

TRONDHEIM_OSLO	9/11/2020 19:00
OSLO_BERGEN	9/11/2020 7:00
OSLO_BERGEN	9/11/2020 8:55
OSLO_BERGEN	9/11/2020 10:40
OSLO_BERGEN	9/11/2020 13:20
OSLO_BERGEN	9/11/2020 14:55
OSLO_BERGEN	9/11/2020 16:20
OSLO_BERGEN	9/11/2020 17:00
OSLO_BERGEN	9/11/2020 19:00
BERGEN_OSLO	9/11/2020 7:30
BERGEN_OSLO	9/11/2020 8:30
BERGEN_OSLO	9/11/2020 9:20
BERGEN_OSLO	9/11/2020 12:00
BERGEN_OSLO	9/11/2020 15:00
BERGEN_OSLO	9/11/2020 16:20
BERGEN_OSLO	9/11/2020 17:40
BERGEN_OSLO	9/11/2020 18:20
OSLO_TROMSO	9/11/2020 8:30
OSLO_TROMSO	9/11/2020 13:45
OSLO_TROMSO	9/11/2020 19:45
TROMSO_OSLO	9/11/2020 6:30
TROMSO_OSLO	9/11/2020 10:55
TROMSO_OSLO	9/11/2020 21:10
STAVANGER_BERGEN	9/11/2020 6:30
STAVANGER_BERGEN	9/11/2020 17:45
BERGEN_STAVANGER	9/11/2020 7:30
BERGEN_STAVANGER	9/11/2020 16:40
TRONDHEIM_BERGEN	9/11/2020 10:20
BERGEN_TRONDHEIM	9/11/2020 11:45

Scandinavian Airlines (SAS) sample

FLIGHT	DEPARTURE
OSLO_STAVANGER	9/7/2020 7:00
OSLO_STAVANGER	9/7/2020 8:15
OSLO_STAVANGER	9/7/2020 9:55
OSLO_STAVANGER	9/7/2020 11:50
OSLO_STAVANGER	9/7/2020 13:30
OSLO_STAVANGER	9/7/2020 15:05
OSLO_STAVANGER	9/7/2020 16:00
OSLO_STAVANGER	9/7/2020 17:15
OSLO_STAVANGER	9/7/2020 19:35
STAVANGER_OSLO	9/7/2020 7:00
STAVANGER_OSLO	9/7/2020 8:25
STAVANGER_OSLO	9/7/2020 9:35
STAVANGER_OSLO	9/7/2020 11:10
STAVANGER_OSLO	9/7/2020 13:05
STAVANGER_OSLO	9/7/2020 16:20
STAVANGER_OSLO	9/7/2020 17:15
STAVANGER_OSLO	9/7/2020 18:30
STAVANGER_OSLO	9/7/2020 20:55
OSLO_TRONDHEIM	9/7/2020 6:05
OSLO_TRONDHEIM	9/7/2020 8:05
OSLO_TRONDHEIM	9/7/2020 10:50
OSLO_TRONDHEIM	9/7/2020 13:10
OSLO_TRONDHEIM	9/7/2020 14:35
OSLO_TRONDHEIM	9/7/2020 15:35
OSLO_TRONDHEIM	9/7/2020 17:20
OSLO_TRONDHEIM	9/7/2020 18:30
OSLO_TRONDHEIM	9/7/2020 20:45
TRONDHEIM_OSLO	9/7/2020 7:00
TRONDHEIM_OSLO	9/7/2020 8:20
TRONDHEIM_OSLO	9/7/2020 9:30
TRONDHEIM_OSLO	9/7/2020 12:25
TRONDHEIM_OSLO	9/7/2020 14:30
TRONDHEIM_OSLO	9/7/2020 15:55
TRONDHEIM_OSLO	9/7/2020 16:55
TRONDHEIM_OSLO	9/7/2020 18:40
TRONDHEIM_OSLO	9/7/2020 19:55
OSLO_BERGEN	9/7/2020 7:05
OSLO_BERGEN	9/7/2020 8:00
OSLO_BERGEN	9/7/2020 10:10

FLIGHT	DEPARTURE
OSLO_STAVANGER	9/11/2020 7:00
OSLO_STAVANGER	9/11/2020 8:15
OSLO_STAVANGER	9/11/2020 9:55
OSLO_STAVANGER	9/11/2020 11:50
OSLO_STAVANGER	9/11/2020 13:30
OSLO_STAVANGER	9/11/2020 15:05
OSLO_STAVANGER	9/11/2020 16:00
OSLO_STAVANGER	9/11/2020 17:15
OSLO_STAVANGER	9/11/2020 19:35
OSLO_STAVANGER	9/11/2020 21:50
STAVANGER_OSLO	9/11/2020 7:00
STAVANGER_OSLO	9/11/2020 8:25
STAVANGER_OSLO	9/11/2020 9:35
STAVANGER_OSLO	9/11/2020 11:10
STAVANGER_OSLO	9/11/2020 13:05
STAVANGER_OSLO	9/11/2020 16:20
STAVANGER_OSLO	9/11/2020 17:15
STAVANGER_OSLO	9/11/2020 18:30
STAVANGER_OSLO	9/11/2020 20:55
OSLO_TRONDHEIM	9/11/2020 6:55
OSLO_TRONDHEIM	9/11/2020 8:05
OSLO_TRONDHEIM	9/11/2020 10:50
OSLO_TRONDHEIM	9/11/2020 13:10
OSLO_TRONDHEIM	9/11/2020 14:35
OSLO_TRONDHEIM	9/11/2020 15:35
OSLO_TRONDHEIM	9/11/2020 17:20
OSLO_TRONDHEIM	9/11/2020 18:30
TRONDHEIM_OSLO	9/11/2020 7:00
TRONDHEIM_OSLO	9/11/2020 8:20
TRONDHEIM_OSLO	9/11/2020 9:30
TRONDHEIM_OSLO	9/11/2020 12:25
TRONDHEIM_OSLO	9/11/2020 14:30
TRONDHEIM_OSLO	9/11/2020 15:55
TRONDHEIM_OSLO	9/11/2020 16:55
TRONDHEIM_OSLO	9/11/2020 18:40
TRONDHEIM_OSLO	9/11/2020 19:55
OSLO_BERGEN	9/11/2020 7:05
OSLO_BERGEN	9/11/2020 8:00
OSLO_BERGEN	9/11/2020 10:10

OSLO_BERGEN	9/7/2020 12:55
OSLO_BERGEN	9/7/2020 14:30
OSLO_BERGEN	9/7/2020 15:30
OSLO_BERGEN	9/7/2020 17:00
OSLO_BERGEN	9/7/2020 18:45
OSLO_BERGEN	9/7/2020 20:40
BERGEN_OSLO	9/7/2020 7:00
BERGEN_OSLO	9/7/2020 8:25
BERGEN_OSLO	9/7/2020 9:20
BERGEN_OSLO	9/7/2020 11:25
BERGEN_OSLO	9/7/2020 14:10
BERGEN_OSLO	9/7/2020 15:45
BERGEN_OSLO	9/7/2020 16:45
BERGEN_OSLO	9/7/2020 18:20
BERGEN_OSLO	9/7/2020 20:20
OSLO_TROMSO	9/7/2020 8:10
OSLO_TROMSO	9/7/2020 9:55
OSLO_TROMSO	9/7/2020 12:05
OSLO_TROMSO	9/7/2020 16:15
OSLO_TROMSO	9/7/2020 17:45
OSLO_TROMSO	9/7/2020 20:15
TROMSO_OSLO	9/7/2020 6:35
TROMSO_OSLO	9/7/2020 10:30
TROMSO_OSLO	9/7/2020 12:15
TROMSO_OSLO	9/7/2020 14:25
TROMSO_OSLO	9/7/2020 18:35
TROMSO_OSLO	9/7/2020 20:05
STAVANGER_TRONDHEIM	9/7/2020 6:00
STAVANGER_TRONDHEIM	9/7/2020 8:00
STAVANGER_TRONDHEIM	9/7/2020 16:10
STAVANGER_TRONDHEIM	9/7/2020 19:35
TRONDHEIM_STAVANGER	9/7/2020 7:00
TRONDHEIM_STAVANGER	9/7/2020 15:25
TRONDHEIM_STAVANGER	9/7/2020 17:45
STAVANGER_BERGEN	9/7/2020 6:35
STAVANGER_BERGEN	9/7/2020 8:55
STAVANGER_BERGEN	9/7/2020 11:10
STAVANGER_BERGEN	9/7/2020 14:35
STAVANGER_BERGEN	9/7/2020 15:35
STAVANGER_BERGEN	9/7/2020 16:10
STAVANGER_BERGEN	9/7/2020 18:50

OSLO_BERGEN	9/11/2020 12:55
OSLO_BERGEN	9/11/2020 14:30
OSLO_BERGEN	9/11/2020 15:30
OSLO_BERGEN	9/11/2020 17:00
OSLO_BERGEN	9/11/2020 18:45
OSLO_BERGEN	9/11/2020 20:40
OSLO_BERGEN	9/11/2020 21:50
BERGEN_OSLO	9/11/2020 7:00
BERGEN_OSLO	9/11/2020 8:25
BERGEN_OSLO	9/11/2020 9:20
BERGEN_OSLO	9/11/2020 11:25
BERGEN_OSLO	9/11/2020 14:10
BERGEN_OSLO	9/11/2020 15:45
BERGEN_OSLO	9/11/2020 16:45
BERGEN_OSLO	9/11/2020 18:20
BERGEN_OSLO	9/11/2020 20:00
OSLO_TROMSO	9/11/2020 8:00
OSLO_TROMSO	9/11/2020 9:55
OSLO_TROMSO	9/11/2020 12:05
OSLO_TROMSO	9/11/2020 16:15
OSLO_TROMSO	9/11/2020 17:45
OSLO_TROMSO	9/11/2020 20:15
OSLO_TROMSO	9/11/2020 22:15
TROMSO_OSLO	9/11/2020 6:35
TROMSO_OSLO	9/11/2020 10:20
TROMSO_OSLO	9/11/2020 12:15
TROMSO_OSLO	9/11/2020 14:25
TROMSO_OSLO	9/11/2020 18:35
TROMSO_OSLO	9/11/2020 20:05
STAVANGER_TRONDHEIM	9/11/2020 8:00
STAVANGER_TRONDHEIM	9/11/2020 16:10
STAVANGER_TRONDHEIM	9/11/2020 19:35
TRONDHEIM_STAVANGER	9/11/2020 7:00
TRONDHEIM_STAVANGER	9/11/2020 15:25
TRONDHEIM_STAVANGER	9/11/2020 17:45
STAVANGER_BERGEN	9/11/2020 6:35
STAVANGER_BERGEN	9/11/2020 8:55
STAVANGER_BERGEN	9/11/2020 11:10
STAVANGER_BERGEN	9/11/2020 14:35
STAVANGER_BERGEN	9/11/2020 15:35
STAVANGER_BERGEN	9/11/2020 16:40

BERGEN_STAVANGER	9/7/2020 7:55
BERGEN_STAVANGER	9/7/2020 9:55
BERGEN_STAVANGER	9/7/2020 13:35
BERGEN_STAVANGER	9/7/2020 15:40
BERGEN_STAVANGER	9/7/2020 16:35
BERGEN_STAVANGER	9/7/2020 17:50
BERGEN_STAVANGER	9/7/2020 19:50
TRONDHEIM_BERGEN	9/7/2020 6:30
TRONDHEIM_BERGEN	9/7/2020 8:30
TRONDHEIM_BERGEN	9/7/2020 12:10
TRONDHEIM_BERGEN	9/7/2020 14:15
TRONDHEIM_BERGEN	9/7/2020 19:20
BERGEN_TRONDHEIM	9/7/2020 6:55
BERGEN_TRONDHEIM	9/7/2020 9:55
BERGEN_TRONDHEIM	9/7/2020 12:10
BERGEN_TRONDHEIM	9/7/2020 15:35
BERGEN_TRONDHEIM	9/7/2020 17:45

STAVANGER_BERGEN	9/11/2020 18:50
BERGEN_STAVANGER	9/11/2020 7:55
BERGEN_STAVANGER	9/11/2020 9:55
BERGEN_STAVANGER	9/11/2020 13:35
BERGEN_STAVANGER	9/11/2020 15:40
BERGEN_STAVANGER	9/11/2020 16:35
BERGEN_STAVANGER	9/11/2020 17:50
BERGEN_STAVANGER	9/11/2020 19:50
TRONDHEIM_BERGEN	9/11/2020 6:30
TRONDHEIM_BERGEN	9/11/2020 8:30
TRONDHEIM_BERGEN	9/11/2020 12:10
TRONDHEIM_BERGEN	9/11/2020 14:15
TRONDHEIM_BERGEN	9/11/2020 19:20
BERGEN_TRONDHEIM	9/11/2020 6:55
BERGEN_TRONDHEIM	9/11/2020 9:55
BERGEN_TRONDHEIM	9/11/2020 12:10
BERGEN_TRONDHEIM	9/11/2020 15:35
BERGEN_TRONDHEIM	9/11/2020 17:45

RStudio codes

Code for analysis of sample for 07.09.2020

```
rm(list=ls(all=TRUE));
library(tidyverse)
library(xlsx)
library(readxl)
library(readr)
library(lubridate)

###Analysis of flights provided by Norwegian and SAS that departed on 7 September 2020
###The aim is to find statistical evidence that specific factors influence the price of the
airline ticket and find out how

head(norwegian0709)
head(sas0709)

###Due to difficulties faced with uploading large files to R we upload missing columns to
datasets
#includeing some missing columns to SAS dataset
sas0709_1 <- cbind(sas0709, competition_departure, sas_missing_values_0709)

#reordering columns
colnames(sas0709_1)
sas0709_2 <- sas0709_1[, c(11, 9, 1, 10, 2, 3, 4, 5, 6, 7, 8)]

#Including missing columns to Norwegian dataset
norwegian0709_1 <- cbind(norwegian0709, Norwegian_missing_values_0709)

colnames(norwegian0709_1)
norwegian0709_2 <- norwegian0709_1[, c(9, 10, 1, 11, 2, 3, 4, 5, 6, 7, 8)]

#merge Norwegian and SAS datasets (Flights departure: 07.09.2020)
library(plyr)
norwegian_sas_0709 <- rbind(norwegian0709_2, sas0709_2) #new dataset that includes
observations for both Norwegian and SAS

###Creating dummy variables based on flight characteristics
```

```

beginning <- ymd_hms("2020-09-07 08:00:00") #early flights depart before 08:00 AM
norwegian_sas_0709$early <- ifelse(norwegian_sas_0709$DEPARTURE<beginning, 1,0)

end <- ymd_hms("2020-09-07 20:00:00") #late flights depart after 20:00 PM
norwegian_sas_0709$late <- ifelse(norwegian_sas_0709$DEPARTURE>end, 1,0)

middle <- ymd_hms("2020-09-07 16:00:00") #flights which depart after 16:00 PM
norwegian_sas_0709$early_evening <- ifelse(norwegian_sas_0709$DEPARTURE>middle, 1,0)

###Registration details dummy variables###
norwegian_sas_0709$late_reg <- ifelse(norwegian_sas_0709$COLLECTION_TIME == 'EVENING', 1,
                                     ifelse(norwegian_sas_0709$COLLECTION_TIME == 'evening', 1,
                                             0))#registered at the late evening

###Dummy variables based on ticket characteristics###
##luggage
norwegian_sas_0709$lug_1 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'LOWFARE', 0, 1) # 1
unit of luggage is included

norwegian_sas_0709$lug_2 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'LOWFARE', 0,
                                   ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'LOWFARE_PLUS', 0,
                                           ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'SAS_GO_SMART', 0,
                                                  1))) #2 units of luggage are included in other categories

##Rebooking
norwegian_sas_0709$reb_1 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'SAS_PLUS_PRO', 1, 0)
#no rebooking fee but fare difference needs to be paid
norwegian_sas_0709$reb_2 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'FLEX', 1,
                                   ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'SAS_PLUS_FULL_FLEX',
                                           1, 0)) #possible to rebook without additional payments

##Refundability
norwegian_sas_0709$ref_1 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == 'SAS_PLUS_SMART', 1,
0) #only refundable for a fee

```



```

norwegian_sas_0709$ref_2 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == "FLEX", 1,
                                ifelse(norwegian_sas_0709$TICKET_CATEGORY == "SAS_PLUS_PRO", 1,
                                        ifelse(norwegian_sas_0709$TICKET_CATEGORY == "SAS_PLUS_FULL_FLEX",
                                                1, 0))) #refundable without additional payments

##Free seat reservation
norwegian_sas_0709$seat <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == "LOWFARE", 0,
                                ifelse(norwegian_sas_0709$TICKET_CATEGORY == "SAS_GO_SMART", 0, 1))
#Those categories exclude free seat reservation

##Priority boarding
norwegian_sas_0709$priority_1 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == "LOWFARE", 1,
                                       ifelse(norwegian_sas_0709$TICKET_CATEGORY == "LOWFARE_PLUS",
                                               1, 0)) #priority boarding is only available for a fee
norwegian_sas_0709$priority_2 <- ifelse(norwegian_sas_0709$TICKET_CATEGORY == "SAS_GO_SMART",
                                         0,
                                         ifelse(norwegian_sas_0709$TICKET_CATEGORY == "LOWFARE", 0,
                                                 ifelse(norwegian_sas_0709$TICKET_CATEGORY == "LOWFARE_PLUS",
                                                         0, 1))) #all categories except those provide priority boarding

##Removing NA rows from the dataset
#norwegian_sas_0709_1 <- na.omit (norwegian_sas_0709)
complete.cases(norwegian_sas_0709)
which(complete.cases(norwegian_sas_0709))
which(!complete.cases(norwegian_sas_0709))

na_values <- which(!complete.cases(norwegian_sas_0709))
norwegian_sas_0709_1 <- norwegian_sas_0709[-na_values,]

###Creating linear models###
library(plm)
library(grDevices)

#Solving problem with duplicated names
make.names(norwegian_sas_0709, unique=TRUE)
anyDuplicated(row.names(norwegian_sas_0709))

```

```

#Dummy variables related to ticket characteristics may be highly correlated with each other in
the model

#Constructing the correlation matrix for all dummy variables
norwegian_sas_0709.dummy = norwegian_sas_0709[c("early", "late", "early_evening", "late_reg",
"lug_1", "lug_2", "reb_1", "reb_2", "ref_1", "ref_2", "seat", "priority_1", "priority_2",
"norwegian")]

norwegian_sas_0709.dummy.cor = cor(norwegian_sas_0709.dummy)

#correlation:
#0.10 - < 0.30 small
#0.30 - < 0.50 medium
# > = 0.50 large

# = 1 perfect correlation which will lead to multicollinearity in the model. Drop some
variables if this is the case

#Perfect correlation: lug_2 and priority2

#Remove priority2

#The pooled model

#Also remove reb_1 and priority_1 due to problem of singularities

#remove ref_1 to avoid misleading results, as well as lug_1 because this variable is not that
important

#Only leave variables which are more insightful

pm_pooled_0709 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2 ,
                    data = norwegian_sas_0709_1,
                    index = c("UNIT", "COLLECTION_DATE"), model="pooling")

summary(pm_pooled_0709)

#Fixed effects model

pm_fixed_0709 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + MARKET_SHARE +
COMPETITION_DEPARTURE + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2 ,
                    data = norwegian_sas_0709_1,
                    index = c("UNIT", "COLLECTION_DATE"), model="within")

summary(pm_fixed_0709)

```

```

#Random effects modell

pm_random_0709 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2,
                    data = norwegian_sas_0709_1,
                    index = c("UNIT", "COLLECTION_DATE"), model="random")

summary(pm_random_0709)

#Chow test

pooltest(pm_pooled_0709, pm_fixed_0709)

#p-value is very small and smaller than 0,05 significance level. Choose fixed effects model

#Hausman test

phtest(pm_fixed_0709, pm_random_0709)

#p-value 0,8 > 0,05. We finally choose random effects model

#Durbin-Watson test for serial correlation

pdwtest(pm_random_0709)

#p-value is very small and smaller than 0,05 significance level. Evidence that there is serial
correlation in the model

#We have large N and small T dataset

#Breusch-Pagan test for heteroscedasticity

library(lmtest)

bptest(pm_random_0709)

#p-value is very small and smaller than 0,05 significance level. Evidence that there is
heteroscedasticity in the model

#Variance inflation factor, doublecheck for multicollinearity

vif(pm_random_0709)

#No values over 5, so we can keep all the variables in the model

#Choose Generalized least squares estimation due to heteroscedasticity and serial correlation
###Our final general model###

library(nlme)

```

```

pm_gls_0709 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE
                + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2, data =
norwegian_sas_0709_1)

summary(pm_gls_0709)

#The coefficients
library(ggplot2)

###Subset model###
#Comparing the most competitive ticket categories
#We want to now how low-cost company (Norwegian) and network company (SAS) set prices in low-
price segment

norwegian_sas_0709_1$economy_lug_1 <- ifelse(norwegian_sas_0709_1$TICKET_CATEGORY ==
"LOWFARE_PLUS", 1,
                                           ifelse(norwegian_sas_0709_1$TICKET_CATEGORY ==
"SAS_GO_SMART", 1, 0))

#Subset: only economy class tickets which can be compared to each other: Norwegian LowFare+
and SAS Go Smart
economy_competition_0709 <- subset(norwegian_sas_0709_1, economy_lug_1 == 1,
                                select = c("UNIT", "COLLECTION_DATE", "FLIGHT",
"DEPARTURE", "TICKET_CATEGORY", "TICKET_PRICE", "DURATION", "TIME_TO_DEPARTURE",
                                "COLLECTION_TIME", "MARKET_SHARE", "COMPETITION_DEPARTURE",
"TIME_TO_DEPARTURE", "early", "late", "early_evening", "late_reg", "norwegian",
"economy_lug_1"))

economy_model_0709 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE + early + late + early_evening + late_reg, data = economy_competition_0709)

summary(economy_model_0709)

###Model for Norwegian (low-cost airline)###
norwegian_sas_0709_1$norwegian <- ifelse(norwegian_sas_0709_1$TICKET_CATEGORY == "LOWFARE", 1,
                                         ifelse(norwegian_sas_0709_1$TICKET_CATEGORY ==
"LOWFARE_PLUS", 1,
                                         ifelse(norwegian_sas_0709_1$TICKET_CATEGORY == "FLEX", 1, 0)))

```

```

lowcost_0709 <- subset(norwegian_sas_0709_1, norwegian == 1,
                      select = c("UNIT", "COLLECTION_DATE", "FLIGHT",
                                "DEPARTURE", "TICKET_CATEGORY", "TICKET_PRICE", "DURATION", "TIME_TO_DEPARTURE",
                                "COLLECTION_TIME", "MARKET_SHARE", "COMPETITION_DEPARTURE", "TIME_TO_DEPARTURE", "early", "late",
                                "early_evening", "late_reg", "lug_2", "reb_2", "ref_2"))
#The model itself
pm_lowcost_0709 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE
                      + early + late + early_evening + late_reg + ref_2, data = lowcost_0709)
#deleted lug_2 and reb_2 due to singularity
summary(pm_lowcost_0709)

###Model for SAS (network airline)###
network_0709 <- subset(norwegian_sas_0709_1, norwegian == 0,
                      select = c("UNIT", "COLLECTION_DATE", "FLIGHT",
                                "DEPARTURE", "TICKET_CATEGORY", "TICKET_PRICE", "DURATION", "TIME_TO_DEPARTURE",
                                "COLLECTION_TIME", "MARKET_SHARE", "COMPETITION_DEPARTURE",
                                "TIME_TO_DEPARTURE", "early", "late", "early_evening", "late_reg", "lug_2", "reb_2", "ref_2"))

pm_network_0709 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + COMPETITION_DEPARTURE +
MARKET_SHARE
                      + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2, data
= network_0709)

summary(pm_network_0709)

```

Code for analysis of sample for 11.09.2020

```
rm(list=ls(all=TRUE));
library(tidyverse)
library(xlsx)
library(readxl)
library(readr)
library(lubridate)

###Analysis of flights provided by Norwegian and SAS that departed on 11 September 2020
###The aim is to find statistical evidence that specific factors influence the price of the airline ticket and find out
how

#Merging the necessary SAS datasets
sas1109_1 <- cbind(sas1109_Copy, sas_missing_values_1109)

#reordering columns
colnames(sas1109_1)
sas1109_2 <- sas1109_1[, c(7, 8, 1, 9, 2, 3, 4, 5, 6, 10, 11)]

#merge Norwegian and SAS datasets (Flights departure: 11.09.2020)
library(plyr)
norwegian_sas_1109 <- rbind(norwegiandata1109, sas1109_2) #new dataset that includes observations for both
Norwegian and SAS

#Deleting NA-values/rows from the dataset

complete.cases(norwegian_sas_1109)
which(complete.cases(norwegian_sas_1109))
which(!complete.cases(norwegian_sas_1109))

na_values <- which(!complete.cases(norwegian_sas_1109))
```

```

norwegian_sas_1109_1 <- norwegian_sas_1109[-na_values,]

#Add dummy variables to the dataset

###Creating dummy variables based on flight characteristics
beginning <- ymd_hms("2020-09-11 08:00:00") #early flights depart before 08:00 AM
norwegian_sas_1109_1$early <- ifelse(norwegian_sas_1109_1$DEPARTURE<beginning, 1,0)
end <- ymd_hms("2020-09-11 20:00:00") #late flights depart after 20:00 PM
norwegian_sas_1109_1$late <- ifelse(norwegian_sas_1109_1$DEPARTURE>end, 1,0)

middle <- ymd_hms("2020-09-11 16:00:00") #flights which depart after 16:00 PM
norwegian_sas_1109_1$early_evening <- ifelse(norwegian_sas_1109_1$DEPARTURE>middle, 1,0)

###Registration details dummy variables###
norwegian_sas_1109_1$late_reg <- ifelse(norwegian_sas_1109_1$COLLECTION_TIME == 'EVENING', 1,
                                     ifelse(norwegian_sas_1109_1$COLLECTION_TIME == 'evening', 1, 0))#registered at the late
evening

###Dummy variables based on ticket characteristics###
##luggage
#Unnecessary dummy variables were deleted from this code for 11.09.2020

norwegian_sas_1109_1$lug_2 <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == 'LOWFARE', 0,
                                   ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == 'LOWFARE_PLUS', 0,
                                           ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == 'SAS_GO_SMART', 0, 1)))
#2 units of luggage are included in other categories

##Rebooking
norwegian_sas_1109_1$reb_2 <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == 'FLEX', 1,
                                   ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == 'SAS_PLUS_FULL_FLEX', 1, 0))
#possible to rebook without additional payments

```

```

##Refundability
norwegian_sas_1109_1$ref_2 <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "FLEX", 1,
    ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "SAS_PLUS_PRO", 1,
        ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "SAS_PLUS_FULL_FLEX",
1, 0))) #refundable without additional payments

#Creating a variable that distinguishes among SAS and Norwegian flights in the pooled dataset
norwegian_sas_1109_1$norwegian <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "FLEX", 1,
    ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "LOWFARE", 1,
        ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "LOWFARE_PLUS", 1,
0))) #1 - Norwegian flight, 0 - SAS flight

###Creating linear models###
library(plm)
library(grDevices)

#Solving problem with duplicated names
make.names(norwegian_sas_1109_1, unique=TRUE)
anyDuplicated(row.names(norwegian_sas_1109_1))

pm_pooled_0911 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + MARKET_SHARE +
COMPETITION_DEPARTURE + early + late + late_reg + lug_2 + reb_2 + ref_2, data = norwegian_sas_1109_1,
    index = c("UNIT", "COLLECTION_DATE"), model="pooling")
summary(pm_pooled_0911)

#Check for multicollinearity
library(car)
vif(pm_pooled_0911)
#no multicollinearity found

```



```
pm_fixed_0911 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + MARKET_SHARE +  
COMPETITION_DEPARTURE + early + late + late_reg + lug_2 + reb_2 + ref_2, data = norwegian_sas_1109_1,
```

```
index = c("UNIT", "COLLECTION_DATE"), model="within")
```

```
summary(pm_fixed_0911)
```

```
pm_random_0911 = plm(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE + MARKET_SHARE +  
COMPETITION_DEPARTURE + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2, data =  
norwegian_sas_1109_1,
```

```
index = c("UNIT", "COLLECTION_DATE"), model="random")
```

```
summary(pm_random_0911)
```

```
#Chow test
```

```
pooltest(pm_pooled_0911, pm_fixed_0911)
```

```
#selecting fixed effects model
```

```
#Hausman test
```

```
phptest(pm_fixed_0911, pm_random_0911)
```

```
#selecting Random effects model
```

```
#Test for serial correlation
```

```
pdwtest(pm_random_0911)
```

```
#there is a serial correlation
```

```
#Test for heteroskedasticity
```

```
library(lmtest)
```

```
bptest(pm_random_0911)
```

```
#there is heteroscedasticity in the model
```

```
#Test for multicollinearity, VIF
```

```
vif(pm_random_0911)
```

```
#Specifying GLS model
```

```

library(nlme)

###Our final general model###

pm_gls_0911 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE +
COMPETITION_DEPARTURE + MARKET_SHARE
                + early + late + early_evening + late_reg + lug_2 + reb_2 + ref_2, data = norwegian_sas_1109_1)

summary(pm_gls_0911)

#Comparing the most competitive ticket categories
#We want to now how low-cost company (Norwegian) and network company (SAS) set prices in low-price segment
###Subset model###

norwegian_sas_1109_1$economy_lug_1 <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY ==
"LOWFARE_PLUS", 1,
                ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "SAS_GO_SMART", 1, 0))

#Subset: only economy class tickets which can be compared to each other: Norwegian LowFare+ and SAS Go
Smart

economy_competition_1109 <- subset(norwegian_sas_1109_1, economy_lug_1 == 1,
                select = c("UNIT", "COLLECTION_DATE", "FLIGHT",
"DEPARTURE", "TICKET_CATEGORY", "TICKET_PRICE", "DURATION", "TIME_TO_DEPARTURE",
                "COLLECTION_TIME", "MARKET_SHARE", "COMPETITION_DEPARTURE",
"TIME_TO_DEPARTURE", "early", "late", "late_reg", "norwegian", "economy_lug_1", "early_evening"))

economy_model_1109 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE +
COMPETITION_DEPARTURE + MARKET_SHARE + early + late + early + early_evening + late_reg, data =
economy_competition_1109)

summary(economy_model_1109)

###Model for Norwegian###

norwegian_sas_1109_1$norwegian <- ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "LOWFARE", 1,
                ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "LOWFARE_PLUS", 1,
                ifelse(norwegian_sas_1109_1$TICKET_CATEGORY == "FLEX", 1, 0)))

lowcost_1109 <- subset(norwegian_sas_1109_1, norwegian == 1,

```

```
select = c("UNIT","COLLECTION_DATE","FLIGHT",  
"DEPARTURE","TICKET_CATEGORY","TICKET_PRICE","DURATION", "TIME_TO_DEPARTURE",  
"COLLECTION_TIME","MARKET_SHARE","COMPETITION_DEPARTURE",  
"TIME_TO_DEPARTURE","early","late", "early_evening", "late_reg", "lug_2", "reb_2", "ref_2"))
```

```
pm_lowcost_1109 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE +  
COMPETITION_DEPARTURE + MARKET_SHARE
```

```
+ early + late + early_evening + late_reg + ref_2, data = lowcost_1109) #deleted lug_2 and reb_2 due  
to singularity
```

```
summary(pm_lowcost_1109)
```

```
###Model for SAS###
```

```
network_1109 <- subset(norwegian_sas_1109_1, norwegian == 0,
```

```
select = c("UNIT","COLLECTION_DATE","FLIGHT",  
"DEPARTURE","TICKET_CATEGORY","TICKET_PRICE","DURATION", "TIME_TO_DEPARTURE",  
"COLLECTION_TIME","MARKET_SHARE","COMPETITION_DEPARTURE",  
"TIME_TO_DEPARTURE","early","late", "early_evening", "late_reg", "lug_2", "reb_2", "ref_2"))
```

```
pm_network_1109 = gls(TICKET_PRICE ~ DURATION + TIME_TO_DEPARTURE +  
COMPETITION_DEPARTURE + MARKET_SHARE
```

```
+ early + late + early_evening + late_reg + ref_2, data = network_1109)
```

```
summary(pm_network_1109)
```

