Pricing Dynamics of Public Service Obligation Routes vs. Commercial Airline Routes: A Comparative Study in Sweden

Thesis

Alexander Kool
2791116
Supervisor: M. Lijesen

Table of contents

1. Abstract……………………………………………………………………………….. 2
2. Introduction……………………………………………………………………………… 2
3. Literature review………………………………………………………………………… 3
4. Methodology…………………………………………………………………………….. 9
5. Results…………………………………………………………………………………….. 16
6. Conclusion and future research …………………………………………………………26
7. References………………………………………………………………………………… 27
8. Appendix……………………………………………………………………………...…… 29

1. Abstract

This thesis explores the differences in pricing dynamics of Swedish domestic airline routes that operate as a Public Service Obligation and regular commercial airline routes. PSO routes are subsidised air routes to ensure connectivity to remote and often economically less active regions. This research aims to determine whether there is a dynamic pricing structure and to what extent, compared to regular commercial airline flights. This study performs a fixed effects regression analysis on data collected via scraping from kayak.nl. Data of all flights on 29 domestic routes are collected for a period of 4 weeks for 7 departure dates. The data includes key variables such as prices, distance of flight, number of stops, day of the week, days before departure and a variable PSO that indicates the PSO status of a flight. Interaction terms are included to compare the effects on regular flights or PSO flights. Results indicate that PSO flights on average have a lower fare than regular commercial flights and have distinct pricing patterns compared to regular commercial flights. Unlike regular commercial flights, PSO flights have a pricing structure that includes both a maximum average fare and shows elements of a dynamic pricing structure. These results fill in a research gap on Public Service Obligations. Public Service Obligations previously only had been studied for welfare implications. The study concludes that there is indeed a dynamic pricing framework for PSO routes and that it is different compared to those in regular commercial airlines. This research contributes to an understanding of pricing on PSO routes and has implications for policymakers and other stakeholders in the Swedish airline industry.

2. Introduction

The aviation industry plays an important role in the connectivity and accessibility of countries, especially those with geographical features like Sweden. Sweden is the 5th largest country in Europe and has numerous less economically viable remote regions. Like some other countries in the European Union, in Sweden, the concept of public service obligations (PSO) was introduced at the beginning of the 21st century. Routes to more remote areas that need additional economic support via accessibility to major airports and cities in Sweden. These PSO routes are subsidised to ensure that residents have access to reliable air travel services. A major aspect of these PSO routes is the level of fares which are different from those prevailing on normal commercial routes due to their unique circumstances therefore this paper aims to establish the differences in air fares between PSO route flights and regular commercial flights within the Swedish domestic flight network.

In this research, the main research question is how pricing dynamics differ between flights operating under a Public Service Obligation and regular commercial airline flights in the domestic air travel market in Sweden. In a statement on fare levels for PSO routes in the PSO inventory table, based on the situation on 17/02/2023. It states that PSO routes in Sweden only apply a maximum average ticket revenue. It is unknown whether this fare can be dynamic or must be a strict fare. In preliminary research performed by myself at the beginning of March 2024, I noticed that PSO operating flights close to departure had different fares than those flights in May. This indicated that there must be some pricing format for those PSO routes. This statement on fares for PSO routes triggered the research question of what this pricing system would look like. The height of the maximum fare is not stated in this document and is unknown. It is further unknown whether airlines operating the PSO routes have advanced, or even any form of revenue management or pricing dynamics.

Despite the role of connectivity between the remote areas and the economic capitals of Sweden, there is a noticeable gap in the fare differences between regular airline routes and those operating under the rules of a PSO. While existing literature is mainly based on the economic impacts of the existing PSO routes, there is zero to no literature on the pricing side of PSO routes in Sweden. This literature gap caused an analysis of the determinants of airline fares for regular and PSO flights on domestic routes in Sweden.

To fill these research gaps, this thesis makes use of fixed effects regression analysis on real-time airline data, scraped from the booking website Kayak.com. Data on all traceable domestic routes in Sweden are gathered for 4 weeks and 7 departure dates. The methodology of the regressions will be explained in Chapter 4. Before that, Chapter 3 contains a thorough review of the existing literature on airline pricing, its determinants and previous research methods will be revised, added with an overview of the Swedish airline industry and Public Service Obligations. In Chapter 5 the regression results will be presented and the following implications of those results will be discussed. In Chapter 6 the conclusions are presented as this thesis hopes to answer the research questions and contribute to the literature on airfare differences in Sweden and their PSO routes.

3. Literature review

Airline pricing dynamics.

A privatised airline is just like any other private firm. Their goal is to make as much profit as they can. In the airline industry, short-run costs are assumed to be fixed (IATA, 2023, p. 42). Operating costs and flight schedules are fixed in the short run. Leaving their turnover to be their main mechanism to maximise profits. Maximising revenues is a very well-known framework in the airline industry, often referred to as revenue management or yield management. Justin Jandor, a revenue management expert explains revenue management as, “Revenue management is the science behind offering the right product at the right price to the right customer at the right time. By finding ways to evaluate a customer’s willingness to pay, revenue management scientists have developed algorithms to adjust a product’s price based on its demand.” In other words, offering the right ticket price, at the right time, considering availability and demand levels. All airlines have their method of revenue management and their associated pricing strategies. Full-service carriers like SAS and Lufthansa, have very advanced systems that even include forms of pricing discrimination, while smaller domestic airlines might apply a very simple and basic form of dynamic pricing. (OAG, 2023)

Aviation industry Sweden.

Sweden is the 5th largest country in Europe based on square metres. It has a surface of 528.000km2 according to the World Databank. Whilst being the 5th largest country in Europe, the population density is very low in Sweden. Each square km counts 24 inhabitants in Sweden, listing itself number 45 out of the 49 registered countries in Europe. The low density in Sweden in combination with the enormous surface, causes Sweden to have many airports and routes. In Sweden, there are around 40 domestic routes between Swedish airports. Those routes are mainly served by Scandinavian Airlines, Braathens Regional Aviation, Norwegian Airlines, Amapola Flyg, Jonair and a few other code sharing airlines. For Scandinavian Airlines, Stockholm-Arlanda serves as one of their hubs in Scandinavia. Scandinavian Airlines is the flag-carrier airline for Sweden, Denmark and Norway. The firm is partly private and partly owned by the Swedish state. Scandinavian Airlines is one of the leading airlines in terms of market share for domestic routes. Together with Braathens Regional Aviation, they account for more than half of the flights in Sweden. Braathens Regional Aviation itself claims to have a market share of around 30% of Swedish domestic flights with 2.2 million passengers per year (Braathens Regional Aviation). Braathens Regional Aviation has its hub at Stockholm-Bromma Airport. Out of the around 40 domestic airline routes in Sweden, most of them fly from and towards Stockholm. Stockholm is the epicentre of the Swedish air network, as the Stockholm area accounts for more than 10% of the Swedish inhabitants. As both of the main national carriers have their hubs based in Stockholm, the remote towns and cities in Sweden need to be connected to destinations outside of Sweden via Stockholm. One of the reasons for having so many air routes in Sweden is the lack of a large railway network that can compete with the aviation network. The road network between the cities is there, however as Sweden is so large, the value of time for flights and driving is not comparable. Some towns and cities are so isolated and thinly populated that airlines need help to serve on those routes to connect the countryside with the urban areas. On top of that, in the winter period, some roads can be poorly accessible due to large amounts of snow and low temperatures. These thin routes are served by airlines that have opted to serve on the route on a Public Service Obligation. In Sweden, there are currently 12 active so-called PSO routes. A document, published by the European Commission on their website that hands out information on the characteristics of all the PSO routes in Europe, presents the average load factor of the PSO routes in Sweden. None of the routes has an annual load factor that exceeded 40%, for the period 2019 to 2023. One might assume that this would not be drastically different for 2024, indicating that fully booked flights are not very common. This would then assume that lack of availability is not primarily the reason for different fares.

Public Service Obligations.

The first Public Service Obligations were imposed by the European Union when the “third package” of European Union air transport liberalisation came into effect on 1st January 1993 (Hromadka, M., 2017). According to the European Commission, the main purpose of PSOs is to maintain appropriate scheduled air services on routes vital for the economic development of the region they serve. Member States may impose public service obligations on these routes. Therefore, they must respect the conditions and requirements in Articles 16-18 of the Air Services Regulation 1008/2008. An auction will be imposed and airlines can bid for the rights to serve on that PSO route. In case of no bidders, the country may restrict access to the route to a single carrier, and compensate for operational losses resulting from the PSO. Currently there are 12 active PSO routes in Sweden. In the PSO inventory table, a document published by the European Commission on February 17th 2023, one can see that compensation or subsidy for all 12 current PSO routes in Sweden has been granted to the operating airline. In the same documents, the number of operators for each route is stated. In Sweden, PSO routes are served by 2 different airlines, Jonair and Amapola Flyg, in the research a third airline will be in the data, this is because FlexFlight is a charter airline that charters aircraft for both Jonair and Amapola Flyg. All 12 PSO routes are served by just one of those airlines, causing the operating airlines to have a monopoly on each of the serving routes. The Airlines operating the PSO routes in Sweden are allowed to make profits once they have won the auction and have the right to operate on the route. Airlines can submit a bid for operating on a route, a winning bid then gives an airline the authority and power to operate for 4 years. The latest PSO contract was signed in October 2023 and is valid till October 23 2027. The PSO routes that are considered in this thesis are visualised in Figure 1. The profits in these contracts are limited, however, as the state supports the airlines with extra funding to cover costs. Airlines therefore are only able to make so-called ‘reasonable profits’, according to the rules and statements by the European Commission.

The main purpose of PSOs is to maintain appropriate scheduled air services on routes vital for the economic development of the region they serve. Previous literature on the welfare effects of PSO routes concluded that for most of the routes subsidising is not socioeconomically beneficial for the Swedish government (P. Förberg, 2023). This study performed a cost-benefit analysis on 7 PSO routes in Sweden for 2015-2019. He concluded that only 2 or 3 routes in the 2015-2019 period were socioeconomically beneficial with subsidies. One of his conclusions was that in order to make these routes more beneficial, the demand for the flights must grow, to increase revenues. The problem with this is that the routes exist because of low demand but still are requested for a small group of people. His advice was to create bus routes instead of airline routes to connect the remote towns with the big cities in Sweden. Kinene et al (2020) concluded the same, that the actual benefits from the PSO flights are lacking socioeconomically for Sweden.

Although this thesis does not include welfare- benefits and costs of PSO routes. It is important to mention the existence and purposes of PSO routes when diving into the pricing of PSO routes.

Factors affecting pricing dynamics.

Airlines thus use revenue management to sell the right product, at the right price, at the right moment, for the right customer. It looks very simple, but the opposite is true. Airline fares have many determinants that affect the actual prices airlines publish.

One of the determinants is competition. When there is competition between airlines on a certain route, airlines can react to the fares of competitors. Rietveld and Pels (2004), did research on the London-Paris market. Initially, their main purpose of research was to investigate whether low-cost airlines and full-service airlines could be seen as clear competitors. The London-Paris market is a very busy route with many competitors, both full-service and low-cost. They found that full-service carriers did not react to price adjustments by low-cost carriers. However, they did find that airlines lowered their fares when competing carriers increased their fares. Their theory on this was that this would be a signal of market saturation. As airlines do not know the load factors of their competitors, they might react to this via their pricing. Another determinant of pricing dynamics is indeed the frequency and capacity of routes. Competition on routes increases the capacity of routes, Malighetti, Paleari and Redondi (2009) did research on Ryanair flights for a year. They found a negative correlation between the average fare and the share of seats offered by carriers. They indicated that when airlines compete on a route, the average fare decreases. This finding is later backed by Gaggero and Piga (2010), as they investigated flights from Ireland to the UK, they found that when Ryanair entered the market, the published fares for Aer Lingus decreased by 8%. Anderson, W. P., Gong, G., & Lakshmanan, T. R. (2002), concluded for the domestic US market, that fares are higher in a market where a single carrier accounts for a very high share of the total passengers than in competitive markets. Indicating that more flights in monopolistic markets have higher fares than in competitive markets. This finding is interesting, whether this holds for PSO routes in Sweden, which are monopolistic as well.

The effects of competition, capacity and flight frequency on routes, do not apply for all routes in the world and are often dealing with the characteristics of a route. Characteristics of routes and markets are important to keep in mind. To name some characteristics. The length of a flight can influence the fares of a flight, as reported by Malighetti, Paleari and Redondi (2009). They found a positive correlation between the length of a flight and the fares. As well as a positive correlation between frequency on a route and fares. Anderson, W. P., Gong, G., & Lakshmanan, T. R. (2002), stated that if distance increases, the fare increases. This relationship between distance and higher fares in the US domestic market is not to be linear.

One would indeed assume that the longer the distance of a flight is, the higher the operating costs are and the higher ticket prices will be, to cover those costs. Airlines often translate the costs of operation directly into ticket prices, as passengers pay for the extra created costs. Wang, B., O'Sullivan, A., Dray, L., Al Zayat, K., & Schäfer, A. (2017, July) state that on average in all markets across the globe, the costs of fuel are the biggest influencer of fares within the window of operating costs.

Another determinant that defines the level of prices are airport taxes, fees or prices that need to be paid to an airport. For simplicity let's call them taxes. Airport taxes are very common taxes in aviation. Airport taxes are fees that are charged to fund the construction, maintenance, and administration of airports and airway systems. These fees are paid by the airlines that operate at the airports. Airlines pay taxes for several actions they perform like, using the runway, attachment to a terminal or congestion taxes (Investopedia, Hussein 2021). The level of the fee depends on the airport type, airports that are fast-paced and often incur congestion, charge higher taxes than small, not congested airports. Congested airports apply higher taxes, due to the high demand levels. The taxes are paid by airlines, however often the customer pays as the taxes are transferred into the ticket prices of flights. The airport taxes are often charged for the departing airport. With the existence of airport taxes based on several factors, the price of flights can be different for the return flights. In this thesis, tax differences are not included, however, differences in airport taxes would be something that could be viable, and to be added in the future.

One of the main determinants that influences fares for a flight is timing. A well-known strategy by airlines is the discount on fares in advance. Pels and Rietveld (2004), found that for most airlines on the Paris-London route, the fares started to increase between 40 and 21 days before departure. Malighetti, Paleari and Redondi (2009), found that on the Italian market, the lowest pricing point for fares was between 50 and 70 days before departure, on average and increasing after. They found that in the final 10 days before departure, the prices increased sharply. While Gaggero and Piga (2010) concluded that for the routes between Ireland and the UK, airline fares just increased as the departure got closer. Unfortunately, they did not have data for their final 7 days before departure. Grabovskaia, S. (2020), a thesis research on the Norwegian Airline market, found results that on average, fares increased between 8.9 NOK and 19.3 NOK per day, depending on the model used. In euros, this would mean an average fare increase of €0.76 and €1.68 per day, based on the exchange level of June 2024. Based on gathered data for a month before departure.

The time of the day for departure is one of the factors that affect the fare level of flights is the departure time of a flight. Grabovskaia, S. (2020), found that departure time influences the fare prices. Flights that departed after 4 pm had higher fares on average than before 4 pm. In the case of this thesis, most routes do not have more than 1 flight option per day and therefore the timing of the flight is probably already based on demand for the departure time.

Another determinant of timing in the fare levels of flights is the day of the week. One could assume that departing on a certain day of the week is more popular than another day, due to weekend days, work days or ending holidays. Evidence of this is still scarce. The reason might be that the day of the week itself is just an occurrence and that the effects are captured in demand levels. Another reason is that travellers often have a reason to travel. They are bound to their scheduled travel days and do not deviate from that. In this research, a holiday weekend is included, and it would be interesting to see prices of tickets differ on certain days of the week.

Methodological Approaches in Airline Pricing Research.

The methodological approaches in the literature regarding airline pricing vary widely. Researchers use various approaches. Data collection approaches range from web scraping to surveys and historical datasets. The variety of statistical approaches that have been used range from simple OLS regression models to more complex models like time-series or lagged distributed models. The methodology per research often depends on the research question and the availability of data. Regarding airline pricing models, the data used for regression analysis often contains flight data over time. Repeating flight data over time often creates a panel data set, in which a cross-sectional analysis over time can be applied. Panel data fixed- or random effects is a methodology that often is presented in the literature. Just as Gaggero and Piga (2010) did, they used an econometric model to study prices and the market structure. Their dependent variable price has been log transformed to interpret the coefficients of the independent variables more easily. They used a panel dataset with cross-sectional flight data that is observed over time. They formed their panel dataset such that each flight/day combination is one panel and observed on each booking day. The regression model used is a random effect panel data regression. An important feature in this research is that they also control for airline's different revenue management strategies. By including fixed effects with the addition of a dummy interaction term to capture Ryanair's approach to intertemporal pricing.

Malighetti, Paleari and Redondi (2009) used a regression analysis to study the determinants of pricing strategies. They used regression models to isolate the effects of individual variables on pricing decisions. They used a panel dataset which contains data for 12 months for each Ryanair route in Europe.

One can conclude that a panel dataset for flight data is a common type of dataset, and can be used for several regression analyses, depending on the research questions and the exact type of data in the set.

Summary and research gaps.

The literature on airline pricing dynamics reveals diverse methodologies and complex pricing determinants. The key determinants of pricing are timing, competition, airport- airline and flight characteristics. An overview of the Swedish airline industry has been presented and reviewed, the literature on pricing on PSO routes is extremely limited, especially in Sweden. Therefore there are some research gaps in the Swedish airline market regarding pricing, especially on PSO flights. This

thesis hopes to fill in a few of those research gaps. One of those gaps is the gap in the existence of a pricing framework on PSO routes. There is little to no literature on pricing on PSO routes. The existence of a pricing framework for PSO flights is therefore not known. Results from a preliminary investigation would assume that there is a framework, as fares do differ over time. The extent of this framework concerning revenue management practices at airlines serving PSO routes is still unknown. In addition to this, there is a research gap in the temporal pricing of PSO flights and non-PSO flights in Sweden. Furthermore, there is still room for further analysis of some other determinants of pricing in the Swedish airline market.
The research gaps will be filled by descriptive analysis of the dataset and by performing 2 fixed effects regressions. A basic regression analysis will be performed to discuss the average pricing differences between PSO flights and regular flights. A more extensive regression model will dive into the temporal pricing of airline tickets and other determinants in pricing dynamics in the Swedish airline industry, to address the potential differences between PSO flights and non-PSO flights. To identify differences in pricing strategies as the departure dates approach, the regressions will include several explanatory and control variables, such as the distance between origin and destination and indicate the impact of a direct or indirect flight. Flights will be sorted per day day-of-the-week to see whether there are effects on certain days of the week. This will be done with the inclusion of an interaction term to see whether there are similarities or differences between PSO and non-PSO flights on this. A difference in this could mean that on one of the two, airlines apply a more advanced strategy of pricing. The week of departure includes a national holiday and the demand for tickets could differ for certain days of the week due to this.

4. Methodology

Data collection process.

The data that is used for this research is generated via scraping. Scraping is a process in which you retrieve data from a live website or web engine. In this research, ticket prices of domestic flights in Sweden on the website Kayak.nl are scraped. Kayak.nl is a metasearch engine for travel services, including airline flights, hotels, rental cars, and vacation packages. Kayak.nl is a very useful and not-so-well-protected web engine to scrape data from. The decision to use this form of data collection was very simple. Historical data on ticket prices is very limited in European countries, and this research requires daily, accurate, and real-time data. This form of data collection was necessary.

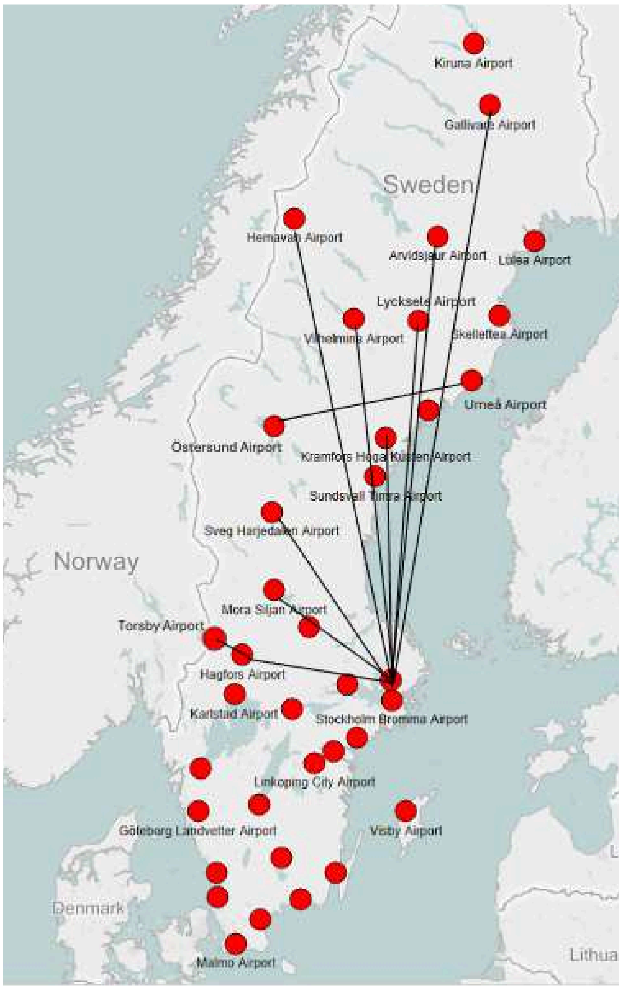
The data that the programme retrieved daily consisted of the following variables. The origin and the destination of the flight, combined with the one-way flight the other way around, and shown with the corresponding IATA city codes. The departure date; scheduled departure time and scheduled total travel time in minutes, with the corresponding airline that performed the flight. In addition, the data consists of whether the route includes the number of stops. To finish the basic dataset the corresponding date of the retrieved data is added in 'scrape date'. The number of routes that are formed with the origins and destinations comes down to 29 domestic routes. These 29 routes are all of the commercial bookable regular domestic air routes in Sweden that were trackable on Kayak.nl, the 29 routes include 11 PSO routes. 1 PSO route is not findable on kayak.nl, while 1 or 2 other PSO routes are not complete or are missing certain days of the week, sometimes. This is mainly because scraping websites is not a 100% reliable method. All the Swedish routes included in the dataset can be found in Table 10 in the appendix**.** The total number of domestic routes in Sweden is somewhere between 35 and 40. Note that in Sweden not all routes operate in all seasons and some of the existing routes are not bookable on Kayak.nl. The dates that were chosen to scrape are May 5th till May 11th. These departure dates have been chosen mainly based on timing, as the process of scraping is very time-consuming. By starting to scrape with a Sunday for the departure dates, this saved me over 2 hours per day in the scraping process. This process started on April 16th and finished on May 11th. The scraping process endured for 25 days straight and took over 2 hours each day. Having data for 25 days straight allowed me to have a large enough dataset with an abundance of observations. It also allowed me to have some spare time after the scraping period for analysis. This dataset expanded daily and therefore it is possible to observe the same observations daily to perform an analysis over time. The database that is retrieved is not the final dataset. In the next section, the raw data will be modified, and cleaned, and new variables will be generated and added.

Dataset preparation.

On May 11th, the data collection process was completed. This resulted in a raw dataset that needed to be prepared for analysis. The dataset formed with this scraping process with R studio was still based on raw data. It contained many outliers, data errors, NAs and unnecessary data. The data cleaning process is mostly performed using the R studio programme. In preparing the dataset for analysis, an important decision was made, by deleting all flights with more than 1 stop. This deleted all, 'unrealistic' flights that any passenger will not book. From this dataset, all flight options contain more than 1 stop. This decision reduced the number of observations from 40,000 to around 36,800 observations. Because this research will focus on the differences between routes that operate under a PSO and those that do not, regarding the pricing dynamics, a binary variable 'PSO', is generated and added to the dataset. This binary variable will later help in the analysis to make distinctions between the routes.

The routes that operate under a PSO are visualised in Figure 1 and are listed in Table 1. With the help of R Studio, the binary variable was easily added to the dataset with the correct flights. In Figure 1 and Table 1, the included PSO routes in this thesis are shown. One can see from the geography of Sweden, that the remote airports that are part of the PSO routes are all above Stockholm. Another interesting variable that is generated after the scraping process, is the variable 'days before departure'. This variable is generated by subtracting the departure date from the scraped date. Days before departure is a well-known variable in the aviation industry. This variable shows the value of ticket prices on the number of days before the flight departs. Adding days before departure is key to filling the literature gap regarding temporal pricing differences and addressing their magnitudes. With the inclusion of the binary variable PSO, the temporal pricing for PSO and non-PSO routes can be compared.

(Figure 1) via planeflighttracker.com



With the availability of the departure date, one could link that date with a certain day of the week. I have generated a variable named, days of the week. This categorical variable contains the day of the week for the associated departure date. This means that May 5th will be classified as a Sunday. May 6th is a Monday, and so forth. Including this variable in the data will allow us to compare the different departure days and whether weekdays are differently priced compared to weekend days. An important reason for this addition is the given that Thursday 9th of May is Ascension Day, which is a celebrated national holiday in Sweden, meaning that many Swedish people only have a 3-day workweek. It is therefore interesting to see whether certain

(*Table 1*) ***List of included Swedish PSO routes***

EU State	Sweden
	Arvidsjaur - Stockholm/Arlanda
	Gällivare - Stockholm/Arlanda
	Hemavan - Stockholm/Arlanda
	Hagfors - Stockholm/ Arlanda
	Kramfors - Stockholm/Arlanda
	Lycksele - Stockholm/Arlanda
	Mora - Stockholm/Arlanda
	Sveg - Stockholm/Arlanda
	Torsby - Stockholm/Arlanda
	Vilhelmina - Stockholm/Arlanda
	Östersund - Umeå
Period of validity of the contract	27 October 2023 - 26 October 2027

days are more expensive than others are, or not. With the inclusion of the binary PSO variable, an interaction term could help to address differences.

In the process of data preparation, I first wanted to include the flight duration. This would be the total scheduled flight time of a flight. I have not included this variable due to the inclusion of another variable that might correlate with flight duration. I have added the variable, flight distance. This is the distance between the origin and destination, and this variable expresses the amount of km between the airports of destination and origin. With the help of airmilescalculator.com and R-studio, I calculated the distance between the origin and destination. One of the limitations that occurred was the attendance of 2 airports in Stockholm. Both are indicated as STO, instead of each having its airport code and location. To solve this issue, I have decided to use the coordinates of an area somewhere between the two airports. The very few kilometres that might be different will be negligible.

For interpretation purposes, I have transformed the variable distance into a log format—newly named log_distance. The same holds for the ticket prices that I have transformed from being expressed in euros, into a log format of the ticket price. log_price. This will later allow us to interpret the coefficients as an elasticity. The elasticity interpretation will be comparable to other literature and allow for implications. Another issue that occurred during the scraping process was the appearance of multiple prices for the same flight. This occurred because two booking options were given on the website. Either an option with a list of different booking sites, or a small discount for booking with the website edreams.com. I believe that this has to do with some sponsor deal. To cover this issue, I have decided that when there are duplicates in observations, to only keep the cheapest option and delete the other prices. This issue with multiple prices for the same flight is addressed in other literature. A thesis report on the validity of web scraping websites reported this issue earlier (Meesters, G. (2021), but a solution has not been given. In this research, to solve this issue, all prices apart from the cheapest option have been deleted.

After cleaning the data, deleting outliers, generating variables, deleting variables and transforming variables. This dataset can be seen as a longitudinal dataset, or panel dataset. These kinds of datasets are characterised by having the same observations observed over time. In this research, the scrape date is considered to be a different variable over time. This setup allows for performing panel data regressions. The main goal of this thesis is to understand and explore how pricing dynamics vary between flights that operate under a Public Service Obligation and regular airline flights that do not, in the domestic air travel market of Sweden. With the format of panel data and a panel data regression, I hope to be able to answer this question and get a better insight into the airline pricing of domestic flights in Sweden.

General models.

The panel dataset that is formed can be regression in multiple ways. In this research panel data fixed effects regressions have been performed. The main reason for this statistical model is that according to the Libguides of Princeton University, “Fixed Effects (FE) explore the relationship between predictor and outcome variables within an entity (country, person, company, etc.). Each entity has its characteristics that may or may not influence the predictor variables. For example, being a male or female could influence the opinion toward a certain issue; or the political system of a particular country could have some effect on trade or GDP, or the business practices of a company may influence its stock price.” In the literature on airlines and fares, often there are fixed effects included to control for route- and or airline characteristics. The dataset for this research includes 29 domestic routes in Sweden, these are presented in Table 10 in the appendix. Most of the flights serving these routes are operated by Scandinavian Airlines, Braathens Regional Airlines, Ryanair, Amapola Flyg, Jonair, Flexflight and Norwegian Airlines. Many different airlines, each having different pricing strategies, performance strategies, cost structures or service levels. Furthermore, airlines make use of different aircraft types. This differs because of route, demand and availability. With all previous arguments, one can formulate that airlines are just not the same. Applying fixed effects on the variable airline controls for the unobserved factors and characteristics that certain airlines have. With this large amount of observations, over 30 different combinations of airlines, including the layover combinations. Controlling for the airlines as a fixed effect is important in this research.

General model. (1)

$$Y_{ijt} = \alpha_j + \beta X_{ijt} + \epsilon_{ijt} \qquad (1)$$

With

- Y_{ijt} as the dependent variable for route *i* and airline *j* at scrapedate *t*.
- α_j as the fixed-effects intercept, capturing the unobserved characteristics over time for airlines
- β as the vector of coefficients for the independent variables X_{ijt}
- X_{it} a vector of independent variables for route *i* and airline *j* and scrapedate *t*
- ϵ_{ijt} error term for route *i*, airline *j* scrapedate *t*

Variables.

In the one-way panel data fixed effect regressions, the dependent variable is the ticket price. The scraped price is the price of a one-way ticket in euros between an origin and a destination. The ticket price has been transformed into a log of the ticket price and is added into an extra column. This is for interpretation purposes only, a log-transformed dependent variable allows us to interpret the coefficients as a percentage effect of the independent variables on the dependent variable. Literature by Gaggero and Piga (2010) used a similar regression and approach. In the panel dataset, the origin and destination each have their column, together with the departure time, departure date and the variable airline. The combination of all those values of variables creates the uniqueness of the flights. One of the independent variables in the regression is the number of connections a flight has. Whether a flight has a stop or not. As previously mentioned, all flights with more than 1 stop have been eliminated from the dataset for 'realistic' purposes. The values in the column ConnectionNumber are indicated with either 'direct' or '1 stop'. One of the key independent variables in this dataset is the binary variable PSO. This variable is very crucial to capture the effects of a flight operating under PSO or not. This variable is binary and thus is either 0 or 1. This makes this variable very useful in an interaction term with other independent variables. Using a binary interaction term will make it much easier to interpret and analyse the independent variables for PSO and non-PSO flights. An interaction term is therefore formed with the independent variable: days before departure. This variable is expressed in whole days. Creating this interaction term, by multiplying the variable PSO with the independent variable, days before departure, will help to analyse the coefficients and show the differences for PSO and non-PSO flights and their days before departure reflecting on the ticket prices. The regressions will include the variable days before departure and additionally in the interaction term with PSO.

Another independent variable that is included in the regressions is the variable day of the week. In the regression there will be 6 coefficients for the days of the week visible, one of them will be omitted from the regression to serve as the reference day. Including all 7 days of the week would violate the assumption of perfect multicollinearity. In the regression, Friday is assigned to be the reference day. Friday, after some trial and error with the reference days, comes out the cleanest for interpretation. This is the day with the cheapest flights compared to the other days, which makes it easy to interpret as the coefficients now are just a comparison of how much more expensive they are compared to Friday. The days of the week can also form an interaction term with PSO and help in the analysis of PSO flights and the day-of-week effect. I have included both interaction terms in the regression. All relevant variables are in Table 2.

(Table 2) ***Variables***

Variable	name	Unit	Information
Price	log_price	log	log of price in euros
Airline	airline	categorial	Airline name
Origin	Origin	City / airport code	airport / city code
Destination	Destination	City / airport code	airport / city code
Departure date	DepartureDate	date	date of departure
Scrape date	ScrapeDate	date	date of scraping
Connection Number	ConnectionNumber	direct or 1-stop	Number of stops
Days before departure	days_before_departure	days	Total of days before the departure
PSO	PSO	binary	either 0 or 1
Day of the week	days_of_week	categorial	day of the week
Distance	log_distance	log	log of distance in km

Regression model specification.

Fixed effects regression model. (2)

$$log\ price_{ijt} = \beta_1 PSO_{ijt} + \beta_2 days\ before\ departure_{ijt} + \beta_3 log\ distance_{ijt} + \beta_4 days\ of\ week_{ijt} + \beta_5 ConnectionNumber_{ijt}$$

$$+ \beta_6 (PSO * days\ before\ departure) + \beta_7 (PSO * days\ of\ week) + \alpha_j + \epsilon_{ijt} \ (3)$$

The regression model (2), will give out coefficients that will be interpreted in the results section in Chapter 5. In this regression model, the dependent variable is the log of the price of a ticket. The independent variables that are included, are explanatory variables or control variables. One of the explanatory variables is the important binary variable PSO. The independent variable days before departure is included to explain the effects of temporal pricing. It reflects the ticket price of a flight, at a certain number of days before departure. The coefficient that will form from this regression, will give the effect on a ticket price for each additional day before departure. This coefficient will be a log-level interpretation, and will therefore be expressed as a percentage in pricing, for each day extra before departure. The interaction term days before departure and PSO, will then give out the interaction effect of PSO and days before departure. As PSO is a binary variable, the total effect can be calculated by hand later in the regression results, to show the differences between PSO flights and regular commercial flights on temporal pricing. The independent variable of distance is a log-transformed variable that controls for the distance between origin and destination. The coefficient in the regression will be an elasticity that indicates the effect in percentages of an additional percentage in distance, on the price. The inclusion of the variable days of the week will give out 6 different coefficients, a coefficient for each day of the week. This coefficient will show the differences in flight fares for non-PSO flights on a certain day in the week compared to the non-PSO flight fares on the reference day. Again the dependent variable is log-transformed. The effects will be reflected in percentages compared to the fares on the reference day, Friday. With this variable, an interaction term with PSO is included as well. The addition of this interaction term will allow us to compare PSO flight fares for certain days of the week with PSO flight fares for the reference day. The coefficients of the interaction days of the week and PSO will be the difference in fares on a certain day of the week for PSO flights, compared to the non-PSO flights on Friday. This can also be reflected in percentages. Again just like for days before departure, the total effect for days before departure on PSO flights can be calculated by hand, by adding both coefficients up. The inclusion of NumberofConnections is to control for the differences between direct and those with 1 stop. The coefficient of this variable will show the difference in average fares for both PSO flights and non-PSO flights in percentages, while controlling for all the other independent variables. The coefficient for variable PSO in this regression requires a very careful interpretation. It is the difference in fares between PSO and non-PSO flights on the reference day when departure date =0. This is because of the addition of the interaction terms. It reflects the effect of the fares on the reference day, which is omitted from the regression results. And at days before departure=0, as that is the 'base' to compare to for days before departure. The coefficient PSO will show the difference between PSO and non-PSO fares on the day of departure at the departure date.

Further, the model specification includes a_j airline-specific fixed effects. Airline-specific fixed effects are there to control for the heterogeneity across airlines, capturing characteristic factors that are constant over time. ϵ_{ijt} is the error term, capturing unobservable factors affecting flight prices. In the regression model that is used in this research, the independent variables ConnectionNumber and distance are mainly included as control variables to control for certain flight characteristics. The extensive interpretation and results of all coefficients of the variables will be discussed in the results section.

5. Results.

Descriptive Statistics.

This part of the research presents the descriptive statistics of the full dataset, the regressions and the key variables. The summary of the dataset is available in Table 3, the dataset contains 36889 observations and is distributed over 25 days of collection. Those observations have been retrieved from 29 routes in Sweden, of which 11 are PSO routes. From Table 3, it is found that 11.5% of all observed flights is a PSO flight, as the PSO variable has a mean of 0.1153. The data summary statistics furthermore tell us that the average ticket price in this dataset is €185.5, and 5.078 in log

units of the price. This indicates that over the period of retrieving data, the average booking price for a domestic flight in Sweden that departs between May 5th and May 11th was 185.5 euros.
The data used in this thesis tells us that Scandinavian Airlines and Braathens Regional Aviation indeed are the 2 largest airlines in Sweden. Together they account for over 60% of all flights in Sweden. The other flights are performed by either Jon Air, Amapola Flyg, Norwegian Airlines, Flixbus or other airlines. Braathens Regional Aviation mentioned on their website that they have around 30% market share in the Swedish domestic market, our results do match their statement.
It is important to notice that there are multiple airlines that have market shares. For this reason there are fixed effects in this model on the variable airline. The fixed effects capture the unobserved characteristics of an airline. The mean values in the summary of the dataset are averages and can only descriptively give means to values. The fixed effects estimates in the regression model are controlled for the other independent variables in the regression models and will give coefficients to interpret. The day of the week with the most observations and thus flights, is Wednesday, with 6544 observations. And the day with the least observations is Saturday, with only 2727 observations. Saturday has the least amount of observations because PSO routes are not active on Saturdays, as well as some more routes being inactive on Saturdays. Furthermore, the average distance for all observations in the dataset is 491 kilometres. This is measured in kilometres in a straight line between the coordinates of the airports of the origin and the destination. The variable distance does not consider the distance from the origin to a stop and to the destination, thus this distance only serves as an indication of the average distance between the airports of routes in this dataset.

Summary Statistics Data variables. *(Table 3)*

	price (€)	**log_price**	**PSO**	**distance (km)**	**log_distance**	**Airline**	**observations**
Min.	11.0	2.398	0.00	222.6	5.405	Scandinavian Airlines	13593
1st Qu.	116.0	4.754	0.00	360.7	5.888	Braathens Regional Aviation	10449
Median	162.0	5.088	0.00	407.6	6.010	Amapola Flyg.	2292
Mean	185.5	5.078	0.1153	491.1	6.010	Norwegian	1923
3rd Qu.	243.0	5.493	0.00	532.2	6.277	FlexFlight	1720
Max.	1738.0	7.460	1.00	1016.6	6.924	others	5956

The average fares of PSO and non-PSO flights on days before departure 19 and 0 are presented in Table 4. At 19 days before departure, each flight has the same amount of observations. The observations for all flights started on a certain date, causing the final departure dates to have more observations and thus days before departure. This is reported as a limitation of the regression model.

The average fare price for a PSO flight, 19 days before departure is €111.30. Whilst this is €129.23, on the day of departure. For non-PSO flights, the average fare is €164.18 19 days before departure compared to €267.62 on the day of departure. Note that these are averages and have not been controlled for flight-specific characteristics, such as distance and airlines.

	PSO	**non-PSO**
Days Before Departure = 0	€129.23	€267.62
Days Before Departure = 19	€111.30	€164.18

(Table 4)

Diagnostic tests.

As in many researches, there is a possibility for bias and some limitations. To minimise limitations and biases in statistical models, several statistical tests can be performed.

In a panel data regression analysis, 3 common statistical approaches can be performed. The random effects regression, the fixed effects regression and the OLS-pooled regression. I have done a Hausman-Wu test to test whether a fixed effects regression is the appropriate regression to use in this research. The results from the Hausman test in Table 5 show that there is no significant difference between the coefficients estimated by the fixed effects and random effects models. The p-value of 0.1916 in the Hausman test suggests that we failed to reject the null hypothesis that the random effects model is consistent with the data. This result indicates that both options are suitable for this data and research. The Hausman test is only a guideline and does not immediately cause the other option to be unsuitable. The reason for using a fixed effect regression instead of random effects regression is that there are indications of unobserved time-invariant factors with the variable airline. The fixed effects model is more suitable for concerns about unobserved individual heterogeneity. In this research, the airline-specific characteristics cause unobserved individual heterogeneity, as all flights are considered to be unique and individual.

Hausman - Wu Test *(Table 5)*

Test Statistic Chi-Square	Degrees of freedom (df)	p-value	Result
13.609	**10**	**0.1916**	**not significant**

A Breusch Pagan test is also performed on the regression model. A Breusch-Pagan test tests for potential heteroscedasticity. When there is heteroscedasticity, the variance of the residuals has a large scatter and the standard errors are biased which can lead to false conclusions. The results of the Breusch-Pagan test in Table 6, indicate that there is indeed heteroscedasticity in the regression model. The p-value of the Breusch Pagan test is lower than 0.05, therefore we reject the null hypothesis and there is indeed heteroscedasticity in the regression. The regressions in this thesis thus have been performed with robust standard errors to account for heteroscedasticity.

Breusch- Pagan test *(Table 6)*

	df	p-value	BP	Result
Fixed effects Regression model	16	2.2e-16	300.23	heteroscedasticity

Regression Results.

In this part, the regression results are presented. The methodology has described the method and the potential interpretations of coefficients from the regression model. The model has been conducted with the dataset that has been formed by scraping published fare data of 29 airline routes in Sweden, Table 10 in the appendix. The regression included 36889 observations. The one-way fixed effect regression has been performed in R studio, and controlled for heteroscedasticity by using robust standard errors. The fixed effects are on the variable airline, to control for the unobserved characteristics of airlines. The R-squared in this regression is 0.43586. The value of R squared explains that around 43.5% of the variance of the log price is explained by this regression model.

Fixed effects regression model. (2)

$$log\ price_{ijt} = \beta_1 PSO_{ijt} + \beta_2 days\ before\ departure_{ijt} + \beta_3 log\ distance_{ijt} + \beta_4 days\ of\ week_{ijt} + \beta_5 ConnectionNumber_{ijt} + \beta_6 (PSO * days\ before\ departure) + \beta_7 (PSO * days\ of\ week) + \alpha_j + \epsilon_{ijt} \quad (2)$$

The regression results for model (2) are presented in Table 7. The coefficient for PSO β_1 is found to be statistically significant at a 0.1% level. β_1 represents the ticket price at days before departure=0, on the reference day, Friday, for PSO flights compared to non-PSO flights. In other words, the average ticket price for a PSO flight on the day of departure on the reference day is compared to the price for a regular flight. This comparison can be expressed as a percentage, fares for PSO flights are 17.49% lower compared to non-PSO flights when days before departure = 0 on the reference day.

Fixed effects Regression Results with Robust Standard Errors (2)

	Dependent variable: log_price
PSO	-0.193***
	(0.046)
days_before_departure	-0.022***
	(0.001)
log_distance	0.170***
	(0.039)
ConnectionNumberdirect	-0.505***
	(0.045)
days_of_weekMonday	0.155***
	(0.036)
days_of_weekTuesday	0.162***
	(0.056)
days_of_weekWednesday	0.261***
	(0.051)
days_of_weekThursday	0.186***
	(0.013)
days_of_weekSaturday	0.144***
	(0.030)
days_of_weekSunday	0.195***
	(0.022)
PSO:days_of_weekMonday	-0.148***
	(0.046)
PSO:days_of_weekTuesday	-0.117*
	(0.062)
PSO:days_of_weekWednesday	-0.165***
	(0.055)
PSO:days_of_weekThursday	-0.054***
	(0.020)
PSO:days_of_weekSunday	-0.021
	(0.072)
PSO:days_before_departure	0.017***
	(0.001)
Observations	36,889
R2	0.436
Adjusted R2	0.435
F Statistic	1,777.557*** (df = 16; 36812)
Note:	*p<0.1; **p<0.05; ***p<0.01

(Table 7)

The coefficient for days before departure β_2 is statistically significant at a 0.1% level. This coefficient indicates that non-PSO flights are 2.19% cheaper for each additional day before departure. This interpretation assumes a linear increase as this is an average difference, expressed in % per day, for at least the final 25 days before departure. All other independent variables are equal. To find the temporal pricing effect for PSO flights, one can calculate the effect by hand. For PSO flights, the pricing difference per additional day before departure is; $\beta_2 + \beta_6 =$ -0.0221904 + 0.0169064 = -0.005284. Indicating a 0.53% decrease in price, each additional day before departure. These findings for PSO routes, are comparable to those for the Norwegian market by Grabovskaia, S. (2020). Assuming an average PSO ticket price of 129 euros would mean that for each additional day before departure the fares would decrease by €0.68. Whilst for non-PSO routes this would be around 4 euros per day.

The coefficient for the log distance, β_3 is statistically significant at a level of 0.1%. Because both the dependent and this independent variable are log transformed, the coefficient can be interpreted as an elasticity. For an additional percentage point of the distance between origin and destination, the average ticket price increases by 0.170%.

The coefficient for ConnectionNumber indicates that direct flights are 39.6% less expensive than flights with 1 stop at a statistical significance level of 0.1%. The reason for not comparing this for PSO and non-PSO routes is that PSO routes are in essence direct routes. Some routes do include a stop, however this is mainly to serve multiple towns with the same aircraft.

The coefficients for days of the week are interesting ones. This variable includes 6 coefficients, one for each day of the week, apart from the reference day. The interpretation of the coefficient for each day of the week is the same. For Monday this is 0.1553260. Indicating that non-PSO flights departing on Monday are 16.80% more expensive, compared to non-PSO flights departing on Friday. For all other days the results can be found in Table 8. All day-of-week coefficients are statistically significant at the 1% significance level. The coefficients of the interaction term PSO * days of the week have the same layout as the days of the week. The reference day is set at Friday, and again the reference day has been omitted from the regression results. The coefficients represent the effect of PSO on days of the week. For Monday, the coefficient is -0.1475187 and statistically significant. This indicates that PSO flights on Mondays are 13.72% cheaper than non-PSO flights on Fridays. The combined effect for PSO flights on Monday is then: -0.1475187 + 0.1553260 = 0.0078073. Indicating that PSO flights on Monday are 0.78% more expensive compared to PSO flights on Friday. Results for the other days of the week are presented in Table 8. All the days are statistically significant enough to be interpreted at a 10% significance level except for PSO flights on Sunday. This coefficient is not statistically significant and therefore cannot be interpreted.

Fare levels per day of the week compared to Friday, in %.

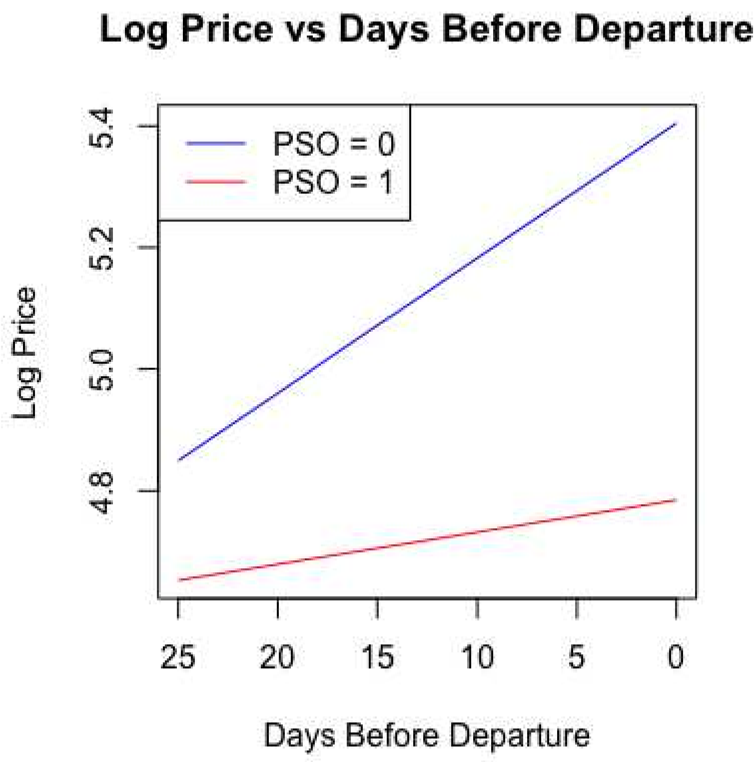
	PSO (%)	**non-PSO (%)**
Monday	0.78	16.80
Tuesday	4.59	17.56
Wednesday	10.07	29.76
Thursday	14.06	20.39
Saturday	NA	15.51
Sunday	19.00***	21.50

*** - not statistically significant at 0.1

(Table 8)

Findings and discussion.

One of the findings in this research is that on average, the prices for a ticket for a flight that operates under a public service obligation are lower than those for non-public service obligation flights. This finding can be seen in Figure 2 and in Table 7. The significant negative coefficient PSO, indicates the finding of a different pricing level for PSO flights compared to regular commercial flights in the Swedish airline market. These findings fill in the research gap that is established in the literature review that PSO operating flights have a pricing framework and that it is different compared to regular commercial flights. In addition to this establishment of a PSO pricing framework, is that the temporal pricing pattern for PSO flights is slightly different compared to that for regular flights. Evidence from this research shows that over time, both PSO and regular flights have increased fares over time. This is in line with evidence from Rietveld and Pels (2004), Gaggero and Piga (2010), Malighetti, Paleari and Redondi (2009) and Grabovskaia, S. (2020). All found evidence that fares increase as departure gets closer. For PSO flights this increase is found to be smaller compared to non-PSO flights. For PSO flights this increase is 0.53% while for non-PSO flights, this is 2.19% for each day closer to departure. A simplified linear pricing curve of the results has been plotted in a graph, in Figure 2. In Figure 2, one can easily see that PSO flights have a less steep increase in fares as the departure date gets closer. The increase rate for PSO routes is comparable to the findings in Norway by Grabovskaia, S. (2020). These findings are done while controlling for distances, number of stops and the number of days before departure the ticket price was found.



(Figure 2, via R Studio)

Figure 2 presents an assumed linear correlation between days before departure and the log price when PSO=0 and PSO=1. The intercepts used in this graph are the averages of the log price values at days before departure =0, for PSO=0 and PSO=1, presented in Table 9. The intercepts, in combination with the coefficients that are retrieved from the regression results of model 2, give a simplified linear pricing curve over time, shown in Figure 2. The log prices are used in this figure for a better visualisation of the differences between PSO and non-PSO.

Average log_price at days before departure=0 for PSO and non-PSO flights.

PSO	days_before_departure	log_price
0	0	5.404580
1	0	4.784920

(Table9)

The coefficient of PSO in the regression results for model 2 in Table 7, represents the difference between the slopes for the fares over time on the day of departure for the reference day, Friday. On the day of departure on Friday, the fares for PSO flights are 17.49% lower than non-PSO flights. From Figure 2, one can see that the fare levels for PSO flights are always cheaper than regular commercial flights. The differences in slope of the graphs indicate that closer to departure, the non-PSO ticket prices have indeed increased much more than PSO ticket prices have. The exact pricing curve over the days before departure was not possible to plot as a graph. The 36,889 observations would take days to complete with the use of a relatively basic computer, and a proper curve would probably not even be visible. To simplify the difference in temporal pricing between PSO and non-PSO flights Figure 2 has been created. It would be interesting to see whether future research can present a more detailed pricing curve. As literature by Malighetti, Paleari and Redondi (2009), found that on Ryanair flights in the final 10 days before departure, the price increased even more. Furthermore, the regulations and rules stated by the European Commission on the pricing of PSO flights argue that fares do have a maximum. Future research can be performed with a different methodology to identify a more exact path of pricing over time on PSO flights, also with the help of an extended dataset. The exact pricing curve might show that regular commercial flights have a steep increase in their final 10 days before departure, compared to the PSO fares that are limited in their fare levels.

A positive coefficient for log distance reveals that flights that cover more distance tend to be more expensive. The found elasticity presents that for each percentage point increase in distance, fares increase by 0.17 percentage points. This aligns with expectations that higher operational costs cause a higher fare. On the other hand, direct connections are priced significantly lower than those with 1 stop, reflecting consumer preference for direct routes and the efficiency of operating non-stop flights. Direct flights are around 39% cheaper than indirect flights. In Sweden, there are not many routes that have both direct and indirect options, as most flights are from and to a hub in Stockholm. From this one might assume that there is not much competition between direct and indirect flights in Sweden, most indirect flights in this dataset are therefore just a summation of two one-way flights, and cause for a higher total fare.

The analysis results reveal interesting patterns related to the days of week pricing. Friday is the cheapest day of the week in pricing for both PSO and non-PSO flights. Each coefficient for the day of the week is positive, indicating that all other days are more expensive compared to the reference day, Friday. The magnitude of the coefficients for days of the week is not the same for PSO and non-PSO flights, indicating that there is no clear pattern of pricing pattern for each day of the week. An explanation for this can be that the data is only based on 1 week, on top of that the chosen week is even a special week that includes a national holiday. Another causation is that PSO routes are just different in demand compared to the bulk of all regular routes in Sweden, the low load-factors on the PSO routes in 2019 indicates this. Whether this hypothesis holds for post-covid periods, is for further research in the future. To conclude for the days of the week is that the evidence that Friday is the cheapest day of the week, for both PSO and regular flights compared to other days is quite remarkable. This is not in line with statements from the industry, as according to vliegtickets.com, Fridays and Sundays are on average the most expensive days to fly on. The demand for those days appears to be the highest as people that leave for a holiday, tend to include their weekends in holidays.

However, in this particular week of departure dates, Thursday is Ascension Day and that might declare that Fridays are lower in ticket prices compared to other days, as people already have their day off on Thursdays and Fridays. On top of that, there is barely any incentive for business travel as it is a national holiday. Indicating that if the same measurement and hypothesis would hold, Wednesdays or Thursdays would be more expensive days to fly on. Evidence from the regression results tells us that Wednesday is indeed the most expensive day to depart on. Notable is that some variables, such as Sunday in the day of the week analysis, are found to be insignificant. This insignificance might stem from the relatively uniform impact of these factors across both PSO and regular flights, the lack of substantial variability in the data for these specific variables, or even the lack of PSO data for Sundays.

Potential limitations and biases and future research.

Noted previously, one of the potential limitations of the model (2) regression is the day of the week variable. The data for this regression only consists of departure dates for 1 week. The assumptions of certain days being cheaper or more expensive than others will be on a single-day comparison. On top of that, the outcomes of the coefficients are not route-specific either. Meaning that on certain routes, the ticket prices of a certain day may be different to others and the results may not be applicable to those routes. Future research could add to this thesis by gathering data that includes more than just a single week of departure dates. This will establish a far more extensive and precise effect on all of the variables and coefficients presented in this thesis. This in addition with PSO route specific regressions to really be able to apply policy decisions for the particular routes.

Another limitation is the variable distance. This distance is calculated as a straight line between the origin and destination, therefore the distance does not take the additional distance for a stop into account as it is not visible in the data where a stop is taking place. I have not corrected for this in the data as I believe the airline fixed effects, ConnectionNumber in combination with the distance can control effects already. For future research, the location of the stops could be added and added to the total distance for a flight.

A bias in the regressions may be that there are missing data observations, as the panel dataset is unbalanced. Some flight data, on the day of departure, is missing. The price of flights that leave before 10 am, will not be scraped on the day of departure. These have not been scraped as the scraping process was mainly performed between 8 and 10 am. On top of that, scraping is not a perfect method of data collection, some flights will not always appear on kayak.nl. Furthermore, not all flights have an equal amount of days before departure, as the process of scraping and data collection started on April 16th, for all departure dates and routes. This would mean that flights that depart on May 11th, have 25 days of data. While flights on May 5th would only be observed for 19 scraping dates. This could affect the outcomes regarding the days before departure. One option could be to level out the playing field for all departure dates, and each has 19 days of observations, however as the number of days before departure is already limited, it has been decided not to. This could be solved by including more days before departure and more periods of departure dates.

Ultimately, this study not only enhances our understanding of the pricing mechanisms in the Swedish airline market but also provides evidence of pricing differences between flights and routes that operate under the rules of PSO that are formed by the European Commission. To inform future policy decisions regarding PSO routes, continued research in this area will be essential for optimising pricing strategies to generate more revenues for PSO flights, and thereby possibly reducing subsidies of the Swedish government and increasing overall welfare in Sweden. The inclusion of new pricing strategies for PSO routes would lead to revisitation of research on the welfare effects of PSO routes, such as in Förberg, 2013.

6. Conclusion and future research.

In conclusion, this thesis has filled in a literature gap on the existence of a pricing framework for PSO flights and compared it to those for regular flights in the Swedish airline market. There is evidence that PSO flights are generally cheaper than regular commercial flights and the hypothesis on PSO fares not being static can be accepted. Literature has already extensively covered general airline pricing dynamics and the impact of competition and timing, but there has been minimal focus on these impacts on PSO routes. With the performance of a fixed effects regression, a research gap concerning the temporal pricing of PSO routes has been filled and compared to literature for non-PSO flights. It is observed that PSO flights exhibit less pronounced price increases compared to regular flights in the final 25 days before departure. This could be attributed to the regulation of PSO routes, which may limit the airlines' ability to implement aggressive revenue management techniques that are available for regular commercial airlines. This aspect of temporal pricing dynamics was previously not explored, and this thesis provides an analysis that enhances the understanding of how PSO flights price their fares over time. The regression analysis also underscores the importance of factors such as flight distance, the day of the week and direct versus indirect flights in determining ticket prices. The main research question on whether PSO flights have different pricing dynamics compared to regular flights on domestic routes in Sweden has been answered by presenting different temporal pricing strategies and price levels for PSO flights and regular flights. The inclusion of the control variable days of the week did not lead to a major difference in pattern for fares per day of the week on PSO flights, compared to those on regular routes. For both groups, evidence is presented that Friday is the cheapest day of that week, contrary to some statements from the airline industry. This thesis not only fills in the established research gaps but also provides practical insights for policymakers and other stakeholders in the airline industry. By presenting new literature on the pricing dynamics of PSO routes and comparing them with regular flights, this research offers a framework for understanding the pricing strategies within the regulated context of PSO operations.

7. References.

- Air Miles Calculator. (z.d.). Air Miles Calculator. https://www.airmilescalculator.com/
- Anderson, W. P., Gong, G., & Lakshmanan, T. R. (2002). Geographical variation in cost of air travel: analysis of the domestic airline fares consumer report. Transportation research record, 1788(1), 13-18.
- Braathens Regional Airlines - Sweden's domestic airline. (z.d.). https://www.flygbra.se/en/about-braathens-regional-airlines
- Förberg, P. (2023). Should the Swedish government continue subsidizing unprofitable domestic flight routes?: A cost-benefit analysis of Swedish air PSOs.
- Gaggero, A. A., & Piga, C. A. (2010). Airline competition in the British Isles. Transportation Research Part E: Logistics and Transportation Review, 46(2), 270-279.
- Grabovskaia, S. (2020). Pricing strategies of airline companies. Discovering the reasons behind ticket price variation on Norwegian airline market (Master's thesis, UiT Norges arktiske universitet).
- Hromadka, M. (2017). Definition of public service obligation potential in the new EU member states. Transport Problems, 12.
- Hussain, A. (2021, December 30). Airport Tax: What it is and How it Works. Investopedia. https://www.investopedia.com/terms/a/airport-tax.asp
- IATA. (2006). Airline cost performance. In IATA.org (IATA Economics Briefing No. 5). https://www.iata.org/en/iata-repository/publications/economic-reports/airline-cost-performance/
- Kayak.nl
- Kinene, A., Granberg, T. A., Polishchuk, V., & Rydergren, C. (2020). Decision support for an optimal choice of subsidised routes in air transportation. Journal of Air Transport Management, 82, 101724.
- Malighetti, P., Paleari, S., & Redondi, R. (2009). Pricing strategies of low-cost airlines: The Ryanair case study. Journal of Air Transport Management, 15(4), 195-203.
- Meesters, G. (2021). Synchronising distributed scraping. Open University of the Netherlands, Heerlen.

- Oag. (2023, November 28). The Evolution of Airline Revenue Management: The Impact of Emerging Technologies | Travel Tech | OAG. Future of travel. https://www.oag.com/blog/airline-revenue-management-impact-technologies
- Op welke dag van de week kan ik het goedkoopst vliegen? (z.d.). Vliegtickets.com. https://www.vliegtickets.com/veel_gestelde_vragen/op-welke-dag-van-de-week-kan-ik-het-goedkoopst-vliegen/
- Pels, E., & Rietveld, P. (2004). Airline pricing behaviour in the London–Paris market. Journal of Air Transport Management, 10(4), 277-281.
- PSO inventory table - situation on 17/02/2023. (2023). In European Commission. European Commission. Accessed on 27 June 2024, from https://transport.ec.europa.eu/transport-modes/air/internal-market/public-service-obligations-psos_en
- Public Service Obligations (PSOs). (z.d.). Mobility And Transport. https://transport.ec.europa.eu/transport-modes/air/internal-market/public-service-obligations-psos_en
- Regulation (EC) No 1008/2008 of the European Parliament and of the Council of 24 September 2008 on common rules for the operation of air services in the Community. (2008). In http://data.europa.eu/eli/reg/2008/1008/oj. European Union Law. Accessed in June 2024, van http://data.europa.eu/eli/reg/2008/1008/oj
- Research Guides: Panel Data Using R: Fixed-effects and Random-effects. (z.d.). https://libguides.princeton.edu/R-Panel
- *SWEDEN AIRPORTS MAP | Plane Flight Tracker*. (z.d.). https://www.planeflighttracker.com/2014/05/sweden-airports-map.html
- Wang, B., O'Sullivan, A., Dray, L., Al Zayat, K., & Schäfer, A. (2017, July). Modelling the Pass-Through of Airline Operating Costs on Average Fares in the Global Aviation Market. In 21st annual ATRS World Conference, University of Antwerp. Summary retrieved from http://www. atslab. org/wp-content/uploads/2017/08/Wang_ATRS_248. pdf.
- What is airline revenue management? (2023, 23 maart). PROS. https://pros.com/learn/airlines-blog/what-is-airline-revenue-management
- World Bank Open Data. (z.d.). World Bank Open Data. https://data.worldbank.org/indicator/AG.SRF.TOTL.K2?locations=SE

8. Appendix.

All included air routes in Sweden *(Table 10)*

City Pairs	
STO	**HFS**
STO	**TYF**
STO	**MXX**
STO	**EVG**
STO	**AJR**
STO	**GEV**
STO	**HMV**
STO	**KRF**
STO	**LYC**
STO	**VHM**
OSD	**UME**
STO	SDL
STO	OER
GOT	VBY
GOT	LLA
STO	RNB
STO	HAD
STO	KLR
STO	KRN
STO	MMX
STO	VBY
STO	GOT
STO	UME
STO	SFT
STO	AGH
STO	OSD
STO	LLA
MMX	VBY
GEV	KRF

^PSO routes