ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis

A first analysis of the on-time performance in the airline-industry: low-cost versus high-cost flight carriers

Gijsbert John Zoutendijk
311385
Prof. Dr. H.P.G. Pennings
Rotterdam, 14-12-2011

Abstract

This research examines the difference between the on-time performance of low-cost flight carriers and high-cost flight carriers across Europe. Results by means of different analyses show no difference in the on-time performance of low-cost carriers and high-cost carriers after controlling for several important variables. Additional analysis shows that Ryanair, on average, has a worse on-time performance than the other low-cost carriers and high-cost carriers, if we control for several other variables. Moreover, we find that the on-time performance of Ryanair is worse than the on-time performance of KLM. We also find evidence for scheduling decisions, type of airport, seasonal effects and some weather and country effects.

Table of Content

1. Introduction

A large amount of research about the airline industry is dedicated to competition and pricing in the airline industry. Although research in this field of the airline industry is interesting, the subject of competition and pricing in the airline industry is saturated. For example Borenstein and Rose (1991) examine the differences in prices that airlines charge to different customers. Another study by Hurdle et al. (1989) investigated the impact of concentration and possible entry of airliners on the performance in the airline industry.

This study will not investigate competition or pricing in the airline industry, but an almost totally new subject related to the airline industry, namely on-time performance of low- and high-cost/national flag flight carriers. The terms high-cost flight carrier and national flag flight carrier are used interchangeably. There is almost no current research about on-time performance of different flights carriers, which is very remarkable. Remarkable in the sense that on-time performance can be considered as one of the major components of flight carrier's strategy.

Suziki (2000), Rupp et al. (2001), Mazzeo (2003) and Prince and Simon (2009) are some of the few researchers who studied the on-time performance of flight carriers. Suziki (2000) examined the relationship between on-time performance and airline market share. Rupp et al. (2001) and Mazzeo (2003) studied the relation between competition in the airline industry and on-time performance. Prince and Simon (2009) investigated the effect of multimarket contact on the on-time performance of flight carriers. All studies find significant effects of the various independent variables on on-time performance.

This research differs from previous studies. To be more specific, this research examines the difference in on-time performance between low-cost carriers and high-cost/national flag carriers. Focussing on the difference between low-cost carriers and high-cost carriers, this leads to the following research question:

Is there a difference in on-time performance between low-cost carriers and high-cost carriers, after controlling for some important variables?

First, we examine the difference in on-time performance between all low-cost carriers and high-cost carriers. The second part of the regression analysis takes the flight carrier Ryanair as reference category. It would be interesting to see if the on-time performance of the largest low-cost carrier in Europe (Ryanair) differs from the on-time performance of carriers in its peer group (EasyJet, Vueling etc.) and from the on-time performance of high-cost/national flag carriers (like KLM and Alitialia).

In 1978 the U.S. was the first to deregulate their airline industry with the Airline Deregulation Act. The U.S. government aimed at increased competition in the airline industry by implementing deregulation measures like removing entry barriers and price restrictions. The deregulations completely changed the market. The airline industry in the U.S. experienced a huge growth in passenger numbers. In 1975 the total number of enplanements for the US was 205,062,000 while the total number of enplanements in 1985, only 7 years after the important Airline Deregulation Act, was already doubled to 382,022,000[1]. With enplanements we mean the total number of passengers boarding an aircraft. Also passenger-miles in the US experienced a dramatic growth. The total passenger-miles in air traffic in 1975 was 147.4 billon and ten years later this number was more than tripled to 290.1 billion[2]. After the deregulation in the U.S. airline industry, other countries followed the strategy of the U.S. by deregulating their domestic aviation market. The European Union applied the same approach. In 1987 the European Council adopted the first package of deregulation measures. The ambition was to create one single aviation market. The process of deregulation in Europe was completed in 1997. Airliners were now free to set prices, to choose their own frequency

[1] The evolution of the airline industry by Steven Morrison & Clifford Winston
[2] Data from Statista, which retrieved their data from Civil Aeronautics Board, U.S. Department of Transportation, Bureau of Transportion Statistics, Office of Airline Information, Eno Transportation Foundation Inc.

and capacity of flights and also to determine whether or not to enter or exit a specific flight route (Burghouwt and Huys, 2003).

The more open aviation market led to new competitors for the incumbent airliners in the airline industry. Removing entry barriers and price restrictions led to the rise of other type of carriers. Southwest introduced the concept of a low-cost strategy in the U.S. After the start of the deregulation measures in Europe in 1993, flight carriers with a low-cost strategy also entered the airline industry in Europe. These low-cost carriers started to compete with low prices. Ryanair began the low-cost revolution in Europe (Burghouwt and Huys, 2003). Low-cost carriers like Ryanair and EasyJet experienced high growth rates varying from 15 to 60% per year (Burghouwt and Huys, 2003). Low-cost carriers faced an improvement in multiple performance measures, like their market share, revenues, profits, revenue passenger-miles and total number of passengers. However, another very important performance measure that is important for both flights carriers and passengers is the on-time performance. Flight carriers want to minimize their delays and their turnaround times in order to maximize the number of flights they can operate with each single aircraft on a single day. This optimizing process will lead to more passengers and higher revenues. Therefore the statement 'time is money' really holds for flight carriers. On the other hand it is clear that passengers benefit from good on-time performance of a flight carrier. Information from Ryanair shows that Ryanair wants to have the best customer service performance in its peer group. Ryanair also claimes to have realized higher punctuality than all other carriers in its peer group. Ryanair has achieved this by operating from uncongested airports[3]. However, it's interesting to see if the on-time performance of Ryanair really is better than other airlines when controlling for other variables.

The different analyses of this research indicate that there is no significant difference between on-time performance of low-cost carriers and high-cost carriers after controlling for some important variables, like scheduling

[3] For more information about the strategy of Ryanair, see www.ryanair.com and go to investor relations.

decisions, type of airport and seasonal effects. Additionally, we do find a significantly worse on-time performance of Ryanair compared to on-time performance of the sample with other low-cost carriers and high-cost carriers, when taking into account the effect of other variables. We also find that the on-time performance of Ryanair is significantly worse than the on-time performance of KLM.

This paper is divided into several chapters. Chapter two consists of an industry overview. We will describe and evaluate the airline industry development of both the U.S. and Europe. We will mention some characteristics of the airline industry and identify players that have an important role in the airline industry. This chapter also describes the circumstances that were responsible for the rise of low-cost carriers. In the third chapter the theory will be further explained. This theoretical chapter consists of related literature. No literature is completely related to the subject of this research, due to the fact that the type of flight carrier related to on-time statistics in the airline industry has never been researched before. However, there is still some literature available that gives an indication of the relevance of on-time statistics and that can be used for a comparison of the results of this study. Consecutively, this chapter also focuses on the causes for flight delay and describes some important aspects related to on-time performance. In the final part of this chapter the hypotheses of this research will be addressed. The fourth chapter is about data and methodology. The first part of the chapter starts with information about the dataset of this research. Numerous components are discussed in this part, like the various data providers, defining dependent, independent and control variables evaluating the dataset by descriptive statistics of numerous variables. The second part of the chapter is about methodology. This part will start with testing the data for normality, constant variance, outliers and multicollinearity. After testing data, we will address our research design by defining the four models in this research and also what kind of analyses will be used for testing the four models. Chapter five will address an important section, namely the analysis of results. In this chapter the results of the four models will be discussed. Each of the four models uses a different independent variable. The first model investigates the relationship between the type of carrier and the delay at arrival of a specific type of carrier. The second model examines the difference in performance between Ryanair and all other carriers in the research. This means the performance of Ryanair will be compared to other low-cost flight carriers, but also to high cost/national flag flight carriers. The third model tries to identify the difference in on-time performance between the low-cost flight carrier Ryanair and other low-cost flight carriers. The last model compares the on-time performance of the low-cost carrier Ryanair, to the on-time performance of the national flag carrier KLM. The first part consists of the first results of the research by identifying the possible relationship between each of the four independent variables and the dependent variable by means of two-group mean comparison tests. Consecutively, there is an analysis using a binary logistic regression of the main model (model 1) in order to see if low-cost carriers differ in whether or not having delay at arrival compared to high-cost carriers. The most important part of the results is the analysis of all models using zero-inflated negative binomial model regressions and negative binomial model regressions. The last part of chapter five consists of the discussion of results. This section examines the various hypotheses of this research and relates the results of this research to the results of related literature. The sixth and final chapter of this paper answers the research question. A general conclusion will be presented, followed by some possible strategy implications for flight carriers. Also some limitations of this research will be given. Finally, this paper ends with possibilities and directions for future research about on-time statistics in the airline industry.

2. Industry overview

This chapter will provide a brief overview of the development of the airline industry in the U.S and Europe[4] by illustrating the change in the structure of the airline industry and by identifying major players in the industry. This overview helps to understand why low-cost carriers were able to revolutionise the total industry and, as a result, forced the large incumbent high-cost carriers to renew their strategy in order to compete with these low-cost carriers and other carriers.

2.1 Development of the U.S. airline industry

In the period after World War I, the commercialisation of the U.S. airline industry took off with air transport for the public. State-owned enterprises and private airliners provided the air transport service. However, the demand for this air transport was very uncertain and therefore the risk of operation was too high (Oum et al., 2010). As a result, the U.S. came up with a subsidizing measure for private airmail in 1925, the so-called Kelly Air Mail Act. This act established a competitive bidding system for private airmail carriage. Following revisions provided explicit subsidies by the Post Office. The Post Office could award contracts with payments exceeding anticipated airmail revenues on the routes (Borenstein and Rose, 2007). However, the real regulation of the U.S. airline industry began in 1938 when the Civil Aeronautics Board (CAB) was created. The role of the CAB was to work as a regulator. The CAB regulated the entry, rate levels and structures, subsidies and merger decisions (Borenstein and Rose, 2007). These regulation measures resulted in limited competition and higher prices (Keeler, 1972). Existing airliners got operating authority over their existing markets and effectively no other companies were allowed to enter the market. This slightly changed during World War II, when the CAB authorized entry of local oriented service flight carriers. The CAB could configure airline networks, because the CAB had the control over the entry of flight carriers on certain routes. As a

[4] See Borenstein and Rose (2007) for an extensive overview about the development of the US airline industry. See Burghouwt and Huys (2003) for an extensive overview about the development Europe airline in Europe.

result, carrier networks were optimized to maintain industry stability and minimize subsidies, but they had no necessary connection to cost-minimizing or profit-maximizing design (Borenstein and Rose, 2007). In the early seventies the U.S. airline industry faced an increase in average cost, rather than an increase in the profitability of airliners, due to a decrease in load factors[5]. Load factors dropped below 50 percent. The industry started its transformation in the mid-seventies. 1978 was the start of the first deregulation measures in the U.S. airline industry, when President Carter signed the Airline Deregulation Act. The U.S. was the first to deregulate their domestic flight industry. The CAB was dissolved in 1985. Market decision-making got an essential role. From now on, it was not the government anymore, who determined the entry of new airliners into the industry or the minimum or maximum of fares that airliners were forced to ask. A new industry structure and competitive environment was created, as a result of the Airline Deregulation Act in 1978. Existing firms expanded their operations and networks, while new airliners entered the market. Average prices dropped and the variation in prices increased. A decreased price level and more variation of prices on a specific route can be attributable to the increased competition. Two types of competition became evident. The first one is price competition. Airliners started with price-based competition. Secondly, other airliners chose to focus on the level of differentiation of their service. More competition led to more turbulence in the industry. Airliners faced financial distress, which forced reorganization and exit. Incumbent airlines reacted to the growing competition by expanding and restructuring their networks. The new hub-and-spoke networks replaced the old point-to-point network created by the CAB. Hub-and-spoke networks have cost, demand and competitive advantages over point-to-point networks. The flight options for passengers increase dramatically under hub-and-spoke networks. For example, if there is not enough demand for a non-stop flight route between New York and Amsterdam, it might be a profitable solution to make it an indirect flight with a stopover in London. You can now target passengers that have interest in a flight from New York to London, New York to Amsterdam or London to

[5] Number of seats sold on a flight divided by the total number of available seats on a flight.

Amsterdam, instead of only having passengers for the flight between New York and Amsterdam. As a result, an airline company can now profit by offering passengers more flight options and therefore increase the load factor on flights. More efficiency and lower costs was a necessary condition for incumbent airliners in order to stay competitive. Another method to achieve competitiveness was merger activity in the mid-eighties. However, due to legislation restrictions like antitrust policies, airliners came up with alternative organizational reforms. Large firms for example, started partnerships with small commuter airlines. Other firms vertically integrated smaller commuter airlines, by buying the small airline instead of starting a partnership. In the late nineties, large carriers started alliances with other large carriers. Economic research finds that alliances have a positive effect on value creation for customers. Bamberger et al. (2001) find that alliances that use code-share connecting routes benefit consumers by lower fares and by more traffic on those pairs of cities that are affected by the alliance. Code sharing stands for a practise in which a particular flight receives the designations of two airlines in the computerized reservation systems (CRS's) by travel agents (Bamberger et al., 2001). Airliners can benefit from the alliances with other airliners. One of the possible benefits is increased consumer demand, because airliners in alliances use inter-airline codeshare.[6] The deregulation also had its effect on the service quality provided by the airlines. Deregulation clearly brings some advantages to customers, like lower fares and more options concerning possible flight routes. However, deregulation also had its effect on, for example, the travel time and flight delays. Substantial expansion in flight operation by airliners and limited improvement and increase in infrastructure led to a dramatic increase in congestion. This congestion not only caused increased scheduled time, but also increased delay beyond the scheduled time. According to The Bureau of Transportation Statistics twenty percent of the flights in 1988 had a delay of more than fifteen minutes (based on the

[6] Inter-airline codeshare: This was defined as a commercial agreement between two airlines under which an airline operating a service allows another airline to offer that service to the traveling public under its own flight designator, even though it does not operate the service (Burton and Hanlon, 1995). For more information about strategic alliances see Rhoades and Lush (1997).

scheduled flight arrival). This percentage increased to twenty-seven in 2000, despite longer implemented flight times.

2.2 Development of the airline industry in Europe

Initially, the European aviation industry was based on bilateral regulation. Since World War II the airline industry relied upon the concept of a national government, the national carrier and the national airport. The air service between two individual nations was negotiated on a bilateral basis (Burghouwt and Huys, 2003). The governments of two nations agreed on the number of airports available for access of a single carrier of each nation. In most cases it was the national carrier (flag carrier) of a country that had permission to fly on the routes between two countries. The airline industry in Europe was heavily centred on the national airline and national airports, during the period of bilateral regulation in Europe (Burghouwt and Huys, 2003). In Europe, there were almost exclusively state-owned airline companies. These state-owed airlines heavily relied upon subsidies of the government. New airliners had practically no chance to enter the airline industry. Therefore there was very limited competition in the European airline industry, just like in the U.S. airline industry before 1978. As a result, Europe also had high fares in the airline industry and airlines did not have any interest to improve their efficiency and lower their cost. During the eighties there was increasing interest for the deregulated airline industry in the U.S. Consumers and airliners became more aware of the positive experiences of the U.S. deregulation measures. Consequently, consumers and airliners started to lobby for deregulation of the European airline industry. The first step in the deregulation of the industry was in 1987, when the European Council adopted the first package of deregulation measures. Europe's goal was to create a single European aviation industry, instead of the previous bilateral agreements between two countries. More deregulation measures were implemented in 1990 and 1993 with the second and third package. The third package was very revolutionary, because it was the European Union (EU) that called for fully deregulated international aviation markets. The EU wanted to replace the bilateral agreements by more multilateral open skies agreements (Borenstein and Rose, 2007). The whole deregulation process in Europe was completed in 1997. Carriers were free to access on any flight route they wanted and both control of prices and national ownership restrictions were eliminated since the completion of the deregulation process in Europe. In 2000, the fifteen member states of the European Union formed a single European airline market. Major European airliners reacted in multiple ways on increased competition, for instance by setting up hub-and-spoke networks. Another reaction was the start of global alliances, like the global alliance of SkyTeam in 2000. This is an alliance of major airlines with for example AirFrance, KLM, Alitalia and Delta Airlines. Finally, airliners also adopted other strategies, like managing a more low-cost strategy. These reactions of incumbent airliners are more or less the same as the reaction of incumbent airliners to airline deregulation in the U.S.

2.3 The Rise of Low-Cost Carriers

Before the deregulation measures in the U.S. in 1978, there already were airliners that really focussed on low operating costs. However, these airliners could not benefit by setting lower fares, because they were not allowed to set prices. This changed with the deregulation of the airline industry in the U.S. and Europe. A new type of competition was born. Airliners were now allowed to determine their own fares and started with competition-based pricing. This new type of competition led to the increase of low-cost carriers. Southwest was the first flight carrier in the U.S. in 1967 that implemented a low-cost strategy. Although many other airliners with a low-cost strategy have entered the market since the deregulation of the industry, many of these low-cost airlines failed in the period from 1980 till today. This phenomenon is very striking, because these carriers focus on having low-cost advantages over national flag carriers. Figure 1 shows the domestic market share of Southwest and all other low-cost carriers in the U.S. in the period of 1984 till 2005 (Borenstein and Rose, 2007)[7]. This figure demonstrates that the market share based on domestic revenue passengers-miles of Southwest and all low-cost carriers has increased dramatically since the deregulation of the market in 1978. Low-cost carriers had a market share of almost twenty-five percent in

[7] Figure comes from Borenstein and Rose (2007). Author calculations from DOT (Department of Transportation) Form 41, Schedule P6. Low-Cost carriers defined as Air Tran, America West, ATA, Frontier, Jet Blue, Midway, People Express, PSA, Reno, Southwest, Spirit, Valujet. Market share based on domestic revenue passenger-miles.

2005. Despite of high failure rates of low-cost carriers, the total market share of low-cost carriers is almost continuously growing since 1980. However, it is still Southwest that is by far the largest low-cost carrier in the U.S. with a market share based on domestic revenue passenger-miles of approximately ten percent. The market share of Southwest based on passengers is even higher with sixteen percent (Ito and Lee, 2003). The counterpart of Southwest was Ryanair in Europe. Ryanair started the low-cost revolution in Europe in 1985 (Burghouwt and Huys, 2003). Ryanair started with flights between Ireland and the British Islands in 1985 and now, in 2011, Ryanair is the largest low-cost carrier in Europe, followed by EasyJet. It is not a coincidence that two of the largest low-cost carriers of Europe started their operations on Ireland and the British islands. Especially the UK had an excellent climate for the rise of low-cost carriers. This climate consisted of low labour costs, a huge London market and a soft regulatory environment in comparison with other countries. After the UK market, these low-cost carriers expanded their operations on a more continental level. In 1997, the year of the completion of the deregulation process in the European airline industry, Ryanair was the first low-cost carrier that launched European Routes. In Ryanair's first year of existence (1985) the total number of passengers was 5,000. Twenty-five years later (2010) the total number of passengers has increased to almost 75 million[8]. But what are the elements that make low-cost carriers so successful? First, the business model of low-cost carriers, like Ryanair, is totally different from national flag carriers like KLM. According to Burghouwt and Huys (2003) the business models of low-cost carriers are based on three key elements: low operating costs, simple products and positioning in the market. Low operating costs can be achieved by focussing on personnel costs. For example, in 2000 Ryanair launched Europe's greatest online booking website. Within three months this site processed an amazing amount of 50,000 bookings a week. There are two predominant advantages of having a booking website. The first one is the low personnel costs, because you do not need a significant amount of personnel, like external travel agents for passenger bookings. Secondly, a booking website enhances the accessibility for

[8] See Ryanair.com, for corporate history and passenger numbers

passengers to book a flight. They do not need to go to an external travel agent, due to the fact that passengers can easily make a reservation at home for the flight they want at any time they want. Low operating costs can also be the result of having low airport fees. The airport fees are the fees that flight carriers need to pay to airports in order to have the landing and take-off rights at a specific airport. Airport fees differ significantly for different types of airports. We make a distinction between main airports and secondary airports. Amsterdam Airport Schiphol for example is a main airport. It can be considered as an international orientated hub airport. It is the primary hub and home base for KLM.[9] In contrast to Amsterdam Airport Schiphol, Rotterdam The Hague Airport is a secondary airport. Such secondary airports are mainly processing point-to-point airlines and charge lower airport fees than main airports. Most landing fees that airports charge depend on three factors: the weight of the aircraft, the type of aircraft and the time of arrival and time of departure. However, in practice, the differences between main and secondary airports in additional charges like surcharge per landing and passenger charges landing fees can be substantial. Landing fees are very important for the choice of airport by flight carriers. High landing fees often result in higher ticket prices. This is one of the reasons why low-cost carriers primarily fly to secondary airports instead of main airports. For example, Ryanair does not fly from or to Schiphol Airport, because it would not be profitable to do so. Also low maintenance costs contribute to lower operating costs. In 1990 Ryanair decided to move to a single type aircraft fleet. Since 2008, Ryanair's entire fleet consists of a single type of aircraft[10].

A single aircraft fleet type has the advantage of lower maintenance costs in comparison with a fleet of multiple types of aircrafts due to limited costs of personnel training, purchase and storage of spare parts. Also greater flexibility of using personnel and equipment is an important factor in order to achieve low maintenance costs. For example, one should take into account that a captain of a Boeing 737 is not simply allowed to fly a larger Boeing 747. Therefore a fleet with multiple types of aircrafts requires employees with

---

[9] KLM is part of the AirFrance-KLM group. AirFrance-KLM is one of the largest airlines with revenues of 23.6 billion euro. KLM is the largest national flag carrier of The Netherlands.
[10] Ryanair operate with a fleet of 250 aircraft. The entire fleet consists of Boeing 737-800's.

different levels of education and experience. The final factor that is important for achieving low operating costs is high daily utilization rate of airports. A high daily utilization rate can only be achieved by reducing turnaround time. The turnaround time is the time that an aircraft must stand at the gate. Turnaround time depends on the aircraft ground handling services. These are the services that are necessary for the next flight of the flight carrier, like baggage handling and refuelling, cleaning and boarding the aircraft. Delay at an airport also has influence on the turnaround time. After the deregulation measures in 1978 in the U.S. and in Europe in 1997, demand was growing dramatically. This huge increase in demand led to more congested airports. It is often assumed that main airports have more delays, because of the large amounts of traffic at these airports. Congested airports can be the result of these large amounts of traffic. Delays at main airports cause larger turnaround times for flight carriers that fly on these main airports. Longer turnaround times implicitly mean lower revenues, because an aircraft only makes money when it is in the air and not when it stands on the ground at the gate. A substantial decrease in turnaround time could mean that the same aircraft can make more flights a day. More flights mean more passengers and more revenues for a flight carrier. Lower turnaround times due to faster ground handling and less traffic at secondary airports could also be a reason why low-cost carriers primarily fly on secondary airports.

The second key element of the business model of low-cost carrier is the no-frills concept. The no-frills concept of low-cost carriers stands for the quality of services that low-cost carriers offer to passengers. By scrapping non-essential services, low-cost carriers are able to charge lower fares than high-cost carriers. No-frills or low-cost carriers only provide the essential services on board of an aircraft and not additional luxury like personal multimedia systems. Passengers must pay for additional services like baggage check-in, snacks and drinks. However, the difference in service level of low-cost carriers in comparison with national flag carriers may not be as significant as most people would think. According to Franke (2004), low cost carriers on continental routes are still able to deliver 80 percent of the service quality at 50 percent of the cost of network carriers. However, at intercontinental routes, this difference in service quality level between the two types of carriers would be greater, while the cost difference to provide this service would probably be lower, because network carriers can benefit from bundling demand in a hub airport. Another factor of the no-frill concept is simple price structures. High-cost carriers offer for example multiple types of service classes on flights, while low-cost carriers mainly have one service class (economy class) on flights.

The final key element in the business models of low-cost carriers is their market positioning. Nowadays, there is a huge demand for short, frequent, cheap, reliable and on-time flights and low-cost carriers can satisfy passengers by offering flights with these characteristics. Low-cost carriers like Ryanair try to keep fares as low as possible, fly frequently to the same destination and focus on limited delays. Initially, low-cost carriers mainly had leisure travellers and passengers on board that wanted to save money on their flight. However, large incumbent airliners also lose business-passengers to low-cost carriers. This is because low-cost carriers offer high frequent reliable and on-time short-haul point-to-point flights (Burghouwt and Huys, 2003).

3. Theory

This chapter will identify and review some related literature to the subject of on-time performance in the airline industry. This chapter also focuses on the causes for flight delay and describes some important aspects related to on-time performance. Finally, this chapter addresses the hypotheses that are necessary for answering the research question mentioned earlier in the introduction.

3.1 Related Literature

In contrast to the many studies that are related to the airline industry, there is only a limited amount of research on on-time performance in the airline industry. None of this, to some extent related literature on on-time performance has investigated the difference in on-time performance between low-cost and high-cost carriers. However, these studies do have some interesting and revealing results.

Suziki (2000) investigates the relationship between on-time performance and market share in the U.S. airline industry. The idea behind the relationship between on-time performance and the market share of a flight carrier is that passengers experience and react to on-time performance of a flight carrier. Passengers are probably more likely to switch to another flight carrier when they experienced substantial flight delays in the past with a specific flight carrier.

The study by Suziki (2000) finds that passengers are more likely to switch, once passengers experience flight delays. Additional results show that on-time performance has a moderating effect on the relationship between passenger's flight experience and market share. However, the results of the study by Suziki (2000) cannot be generalized, since this study is based on single-route data.

Another study by Rupp et al. (2001) focuses on the influence of route competition on on-time performance and other factors that flight carriers should take into account while determining the delay of their flights. The main finding of the study by Rupp et al. (2001) is that more competitive routes have worse on-time performance. Additional (control) variables like seasonal effects, airport capacity constraints, the number of scheduled flights, hub originations[11] and prior month's performance also had a significant effect on on-time performance of flight carriers.

The study by Mazzeo (2003) examines more or less the same as Rupp et al. (2001). Both Mazzeo (2003) and Rupp et al. (2001) investigate whether or not the lack of competition affects the on-time performance of flight carriers. However, results of both studies are conflicting. Mazzeo (2003) finds that both the prevalence and duration of flight delays are significantly greater on routes where only one airline provides direct service and additional competition is correlated with better on-time performance. In contrast to Mazzeo (2003), Rupp et al. (2001) find that the more competitive routes have worse on-time performance. Therefore the exact relationship between route competition and on-time performance is not clear yet. Mazzeo (2003) also used some control variables in his research. Control variables, like weather conditions, congestion and scheduling decisions also have a significant effect on flight delays.

One of the most recent papers in this field of research is the paper by Prince and Simon (2009). They examine the relationship between multimarket contact and service quality. Multimarket contact means that a firm competes with its rivals in multiple markets. Multimarket contact is a frequently occurring phenomenon in the airline industry, because flight carriers often compete with the same rivals on different routes. Focusing on this relationship helps to understand how firms may vary on the service quality level in response to changing competitive conditions. On-time performance is used as a proxy for service-quality. The main result of the study by Prince and Simon (2009) is that multimarket contact has a negative effect on the on-time performance of flight carriers and this effect is greater for contacts on more concentrated routes. Prince and Simon (2009) attribute this negative multimarket contact

[11] The variable hub origination equals 1 if a flight of a carrier departed from the hub of that same carrier.

effect on on-time performance to mutual forbearance. This mutual forbearance reduces the incentive of firms to invest in service quality.

3.2 Causes Flight Delay

Besides the research findings of Suziki (2000), Rupp et al. (2001), Mazzeo (2003) and Prince and Simon (2009) mentioned in paragraph 3.1, the Bureau of Transportation Statistics (BTS) in the U.S. also specified some causes for flight delays and poor on-time performance. The BTS identified several factors that contribute to flight delays. It divided the causes of flight delays in five categories[12]: air carrier, extreme weather, National Aviation System (NAS), late-arriving aircraft and security. Table 13 in the appendix shows that flight delays are primarily caused by the flight carrier itself, aircrafts arriving late and by the NAS. Especially, the category air carrier as explanation for delay is interesting, because many airlines argue that flight delays are outside their control. Flight carriers can influence some circumstances, like aircraft cleaning, while other categories of flight delays are not controllable by the flight carrier. Also, Borenstein and Rose (2007) identified some disruptions, like weather conditions, congestion externalities and inefficient infrastructure investments that cause flight delays.

3.3 On-Time Performance

On-time performance is often used as a proxy for the level of service quality of airlines. Service quality is a performance measure of many firms. It is important for both passengers and airliners. It is clear that passengers want to know which flight carriers have the best service performance. On the other hand, flight carriers themselves are also curious about the quality of their service. However, flight carriers are more interested in which factors influence their service quality, so both the process leading to a high service level as well as the outcome of the process is important for flight carriers. Identifying these factors can help flight carriers to optimize their service quality level. Studies of Rupp et al. (2001), Mazzeo (2003) and Prince and Simon (2009) use on-time performance as a proxy for on-time performance in the airline industry.

[12] See appendix table 12 for more information about the five categories of flight delay.

Mazzeo (2003) suggests that on-time performance is an appropriate and commonly used proxy for service quality, because Bowen and Headley (2001) state in their widely cited 'Airline Quality Ratings 2001' that on-time performance is one of the elements for determining an airline's quality rating. Therefore, different researchers use the on-time performance in order to measure the service quality of flight carriers.

However, on-time performance is not only important for measuring the service performance of an airline, but also for measuring financial performance. Of course, on-time performance is more closely related to service performance, though on-time performance could also have a connection with financial performance. The fact that on-time performance and turnaround time are both important for financial performance of flight carriers was already stated in the previous chapter. Fewer delays and shorter turnaround times can improve a firm's financial performance.

3.4 Hypotheses

Based on information and inspiration from current research on on-time performance in the airline industry, we were able to determine some interesting hypotheses for this research. However, it is hard to base the hypotheses on well-studied theories. This is because there is no research that has investigated the difference in on-time performance between low- and high cost-carriers. Hypotheses in this study are rather intuitive. For the first hypothesis, we expect no difference in on-time performance between low cost carriers and high cost carriers after controlling for some important variables. A significant difference after controlling for several variables could indicate that there are other important variables that are related to on-time performance of flight carriers, like management performance of the firm. Still, we do not have any theoretical foundation that gives an indication for a significant difference between low- and high-cost carriers. As a result, we have established the first/main hypothesis:

H1: there is no difference between the on-time performance of low-cost carriers and the on-time performance of high-cost carriers.

The second, third and fourth hypotheses are based on the relative performance of the low-cost carrier Ryanair. Ryanair states that they have realized higher punctuality than all other carriers in its peer group. In 2005, The Civil Aviation Authority (CAA) came up with some punctuality statistics. As a result, Ryanair proudly mentions its superior on-time performance compared to EasyJet and calls Easyjet 'Easy-to-be-late-Jet' in a marketing campaign. In order to check the performance of Ryanair, Ryanair is taken as base category in hypothesis H2, H3 and H4. The second hypothesis is:

H2: there is no difference between the on-time performance of Ryanair and the on-time performance of other low-cost carriers and high-cost carriers.

The third hypothesis is based on the on-time performance of Ryanair compared to the on-time performance of other low-cost carriers in this research.

H3: there is no difference between the on-time performance of Ryanair and the on-time performance of other low-cost carriers.

The fourth and final hypothesis is about the comparison between the on-time performance of Europe's largest low-cost carrier Ryanair and the on-time performance of the well-known national flag carrier KLM.

H4: there is no difference between the on-time performance of Ryanair and the on-time performance of KLM.

4. Data and Methodology

This chapter gives more information about the data providers, dataset, variables definitions, variable descriptive statistics, research methodology and models of this research.

4.1 Data and Variables

This research is based on a completely new dataset. The dataset is specially built and organized for this study. Flight data is provided by FLIGHTSTATS. FLIGHTSTATS is parented by Conducive Technology Corp. It offers different information and solutions for specific customers. Travellers have easy access to a lot of information that is relevant for their air travel. For example, they can easily find and evaluate information about the statistics of a certain flight or flight route. FLIGHTSTATS also offers solutions to airports, airlines, travel agencies etc.[13] Weather data is provided by EUROWEATHER. EUROWEATHER has a large database with different types of weather variables, like visibility, temperature, sky conditions etc.

This study evaluates the on-time performance of different airline companies based on route level data. According to Rupp et al. (2001) an advantage of route level data is the ability to control for route-specific effects in addition to carrier and month-specific effects. The dataset contains three dependent variables. The first one is a binary dependent variable *delay_a_minutes_0/1* and is defined as the delay at arrival of a certain flight in minutes. This variable takes the value of 0 when a flight arrives at the airport within fifteen minutes of scheduled arrival time and takes the value of 1 when it is delayed. This is the case when a flight arrives fifteen minutes or more behind the scheduled time of arrival at the airport.[14] The second dependent variable is *delay_a_minutes(1)*. This is a continuous variable and is defined as the delay in minutes of a flight at arrival. This variable both contains negative values and positive values. Negatives values mean that a flight has no delay and

[13] For more information about FLIGHTSTATS visit http://www.flightstats-inc.com/
[14] According to the Federal Aviation Administration a flight is delayed if it arrives or departures fifteen minutes or later than scheduled arrival or departure time.

positive values mean that a flight has delay at arrival. It should also be noted that we winsorize the negative and positive outliers by means of μ+-2xσ.[15] The third dependent variable *delay_a_minutes (2)* is derived from *delay_a_minutes(1)* and is also a continuous variable. This dependent variable is defined as the delay in minutes of a flight at arrival. In contrast to *delay_a_minutes(1)*, *delay_a_minutes (2)* can only take on values equal to or larger than zero. Negative values are set to zero. Positive outliers are also winsorized by means of μ +2xσ.[16]

The dataset consist of four independent variables (table 1). The first one is *type_flight_carrier*. This variable takes on the value of 0 when the flight carrier is a high cost/national flag flight carrier and takes on the value of 1 when the flight carrier is a low cost flight carrier. The second independent variable is *Ryanair_other_carriers*. The value of this variable is 0 when the flight carrier is a low-cost flight carrier and not Ryanair or when the flight carrier is a high-cost/national flag flight carrier. Consecutively, the value of this independent variable is 1 when the flight carrier is Ryanair. The third independent variable is *Ryanair_other_low_cost_flight_carriers*. All low-cost flight carriers except Ryanair are denoted by the value 0. The variable has the value 1 when the flight carrier is Ryanair. Finally there is the independent variable *Ryanair_KLM*. This variable takes on the value of 0 when the flight carrier is KLM and the value of the variable is 1 when the flight carrier is Ryanair.

Obviously, this research also used several control variables (table 1). These control variables can have effect on on-time performance of flight carriers. Adding control variables to our research will lead to more reliable results. The first control variable is *time_sched_d*. The variable stands for the scheduled time of departure of a specific flight. The variables can only take on discrete values between 0 and 23. For example, when a flight has a scheduled time of departure of 0.15am, than the value of the variable *time_sched_d* is 0 and when a flight has a scheduled time of departure of 11.25pm the value of the

[15] Outliers are winsorized by means of μ+- 2xσ. Values smaller than μ-2xσ are set to μ-2xσ and values larger than μ+2xσ are set to μ+2xσ.
[16] Positive outliers are winsorized by means of μ + 2xσ. Values larger than μ+2xσ are set to μ+2xσ.

variable will be 23. It is argued that flights with a later scheduled departure time are more likely to have flight delays at arrival. For example flights with a scheduled departure time in the evening are more prone to flight delays, because these evening flights are dependent upon the flight delay of previous flights. If for example KLM has multiple flights from Amsterdam Airport Schiphol to Rome Fiumicino on a single day, the probability of flight delay is the lowest for the first flight on that day, while the probability of flight delay is the highest for the last flight. Mazzeo (2003) finds evidence for a significant effect of scheduling decisions.
The variable *type_airport_a* is the type of airport where a specific flight carrier arrives. There are two types of airports: main airports and secondary airports. The value of this variable is 0 when the airport is a secondary airport, while the value is 1 when the airport is a main airport. For example Amsterdam Schiphol Airport (AMS) is a main airport, while Eindhoven Airport (EIN) is a secondary airport. It is expected that flights that arrive at a main airports are more likely to have a delay than flights that arrive at a secondary airport due to negative congestion externalities as a result of the large amounts of traffic at main airports. The variable *date* stands for the month in which a particular flight took place. The value is 0 when the flight took place in December 2010 and the value is 1 when the flight took place in April 2011. With this *date* variable we want to control for seasonal effects. Rupp et al. (2001) already used seasonal effects in order to explain on-time performance of flight carriers. Seasonal effects, like severe weather conditions in wintertime and more leisure travellers during the summer, probably have a negative effect on on-time performance. Therefore our expectation is that flights from December 2010 have a worse on-time performance than flights from April 2011.
Then there is the distance variable, defined by the distance of a specific flight route. The distance of a flight is measured as the kilometres in a straight line between the airport of departure and the airport of arrival. More specific, the log of the distance in kilometres is taken for constructing the variable. According to Rupp et al. (2001) pilots can 'make-up' for the flight delays on the ground at the airport by flying at higher air speed. Therefore we expect that the variable *log_distance* has a positive effect on the on-time performance. Another control variable is *visibility_a_1/2/3/4/5*. *Visibility_a* can

be defined as the visibility at the airport of arrival at the moment a flight arrives. This is a dummy variable, which can take on values of 1,2,3,4, or 5. The values 1/2/3/4/5 represent consecutively the categories very bad, bad, poor, fair and good visibility. This variable *visibility_a_1/2/3/4/5* is a weather variable and though it may partially measure the same as the seasonal effects in the *date* variable it is not exactly the same. Visibility is really a weather specific measure, while seasonal effects not only measure weather specific effects, but also for example take holiday-rush into account. The final control variable is *country_a1/2/3/4*. This variable can be defined as the country of arrival of a specific flight. The variables can take on four values: 1,2,3,4, which consecutively denote The Netherlands, Spain, Italy and Sweden. Significant country dummies could be the result of differences in the general quality of the infrastructure in the airline industry. For example, this variable might give some indication for inefficiencies in the air traffic control system of a country.

Table 1: Definitions of the different dependent, independent and control variables

Variables	
Dependent Variables	
delay_a_minutes_0/1	Discrete variable, defined by the delay in minutes of a flight at arrival, that takes on value 0 when delay is < 15 minutes or 1 when delay ≥ 15 minutes
delay_a_minutes(1)	Continuous variable, defined by the delay in minutes at arrival. Both negative and positive outliers are winsorized by means of μ +- 2xσ.
delay_a_minutes(2)	Continuous variable, defined by the delay in minutes at arrival, Positive outliers are winsorized by means of μ +- 2xσ. Negative values are set to 0.
Independent Variables	
type_flight_carrier	Discrete variable, takes on value 0 when the flight carrier is a high cost/national flag carrier or 1 when the flight carrier is a low-cost carrier
Ryanair_other_carriers	Discrete variable, takes on value 0 when the flight carrier is a low-cost carrier and not Ryanair or when a flight carrier is a high-cost/national flag carrier and 1 when the flight carrier is Ryanair
Ryanair_other_low_cost_flight_ carriers	Discrete variable, takes on value 0 when the flight carrier is a low-cost flight carrier and not Ryanair and 1 when the flight carrier is Ryanair.
Ryanair_KLM	Discrete variable, takes on value 0 when the flight carrier is KLM and 1 when the flight carrier is Ryanair
Control Variables	
time_sched_d	Continuous variable, the variable can take on values between 0 and 23. The values represent the scheduled departure time of a flight.
type_airport_a	Discrete variable, takes on value 0 when the airport of arrival is a secondary airport and 1 when the airport of arrival is a main airport
date	Discrete variable, takes on value 0 when a flight took place in December 2010 or 1 when the flight took place in April 2011
log_distance	Continuous variable, defined as the log of the distance in kilometres of a specific flight. This variable is constructed by measuring the distance in a straight line between two airports on a flight route.
visibility_a1/2/3/4/5	Discrete variable, defined as the visibility at the time of arrival at the airport of arrival. This variable can take on five values: 1,2,3,4,5, which represent respectively very bad, bad, poor, fair and good visibility
country_a1/2/3/4	Discrete variable, defined as the country of arrival of a specific flight. This variables can take on four values: 1/2/3/4, which represent respectively The Netherlands, Spain, Italy and Sweden

The dataset consist of flights from 2010 and 2011 from eight flight carriers, to be more specific Alitalia, EasyJet, KLM, Norwegian Air, Ryanair, SAS, Transavia and Vueling. This research only uses flights on a specific flight route on which there was at least one flight within two weeks. The flight route was deleted from the dataset when there was no flight during one week on this specific route, otherwise this specific flight route had too little observations. The data contains flights between eleven different airports across four countries. From the Netherlands we used the following airports: Amsterdam Schiphol Airport (AMS), Rotterdam The Hague (RTM), Eindhoven Airport (EIN) and Maastricht Aachen Airport (MST). Barcelona Airport (BCN), Barcelona Girona Airport (GRO) and Reus Airport (REU) are the Spanish airports in the dataset. Consecutively, Rome Fiumicino (FCO) and Rome Ciampino Airport (CIA) are the airports in Italy. Finally, there are 2 airports from Sweden: Stockholm-Arlanda Airport (ARN) and Stockholm Skavsta (NYO). The total number of flights arrived in The Netherlands, Spain, Italy and Sweden is consecutively 360, 527, 352 and 235 (appendix, table 15). The number of flights that arrived at a main airport in December 2010 and April 2011 is higher than the number of flights that arrived at a secondary airport in the same periods. Main airports handled 961 flights in total and secondary airports processed a total amount of 513 flights. 533 out of 961 flights to main airports and 233 out of 513 flights to secondary airports were delayed at arrival. Therefore the percentage of flights that have a delay at arrival is higher for main airports (55.46%) than for secondary airports (45.42%)(appendix, table 16). In total there are 35 flight routes in the dataset (appendix, table 14). We examined 289 flights from the month December in 2010. In 2011 the total number of flights in April was 1185. Table 2 shows the descriptive statistics of the variables. This research uses a lot of discrete variables within two categories. For these discrete variables, the mean, standard deviation, min and max do not have value. Therefore, table 2 shows the total number of observations and the total number of observations for the separate categories (0/1, 1/2/3/4/5 and 1/2/3/4) of the variables. For example, the dependent variable *delay_a_minutes_0/1* has a total of 1474 observations and 1119 observations have a value of 0. This means that 1119 out of 1474 flights have a delay of less than 15 minutes. *Visibility_a* consists of five categories (table 1). 43 observations had a value of 2. This means that 43 flights faced bad visibility at the airport of arrival.

Table 2: Descriptive statistics of the variables. The mean, standard deviation (St. Dev.), Min and Max are only displayed for continuous variables. The total number of observations, the total number of observations of a certain value (0 or 1) or the total number of observations for each dummy is displayed for the discrete variables.

Variable	**Obs.**			**Mean**	**St. Dev.**	**Min**	**Max**
	Total	**0**	**1**				
Dependent variables							
delay_a_minutes_0/1	1474	1119	355				
delay_a_minutes	1474			10.77	37.96	-36	433
delay_a_minutes(1)	1474			8.15	24.71	-36	87
delay_a_minutes(2)	1474			12.47	21.41	0	87
Independent variables							
type_flight_carrier	1474						
Ryanair_other_carriers	1474	921	553				
Ryanair_other_low_cost_flight_carriers	868	315	553				
Ryanair_KLM	956	403	553				
Control variables							
time_sched_d	1221			12.77	4.72	5	21
type_airport_a	1474	513	961				
date	1474	289	1185				
log_distance	1474			7.13	0.28	6.70	7.75
		1/2/3/4/5					
visibility_a1/2/3/4/5	1416	5/43/76/201/1091					
		1/2/3/4					
country_a1/2/3/4	1474	360/527/252/235					

Table 17 in the appendix shows some descriptive statistics for the individual flight carriers. From this table we can see that most flights (553) are from Ryanair. Ryanair had an average delay of 6.17 minutes before winsorizing outliers in the dataset. After winsorizing influential outliers the average delay of Ryanair reduced to 4.12 minutes. EasyJet is the flight carrier with the highest average delay before winsorizing outliers (20.39 minutes). KLM is the flight carrier with the highest average delay (14.64 minutes), after winsorizing the outliers. The last column of ‘Mean delay arrival in minutes’ shows the average delay in minutes after winsorizing the outliers and setting negative values to zero. According to these results, KLM has the highest average delay, after winsorizing the positive outliers and setting negative values to zero.

4.2 Methodology

We should first examine the data in order to choose the most appropriate type of regressions. After this data examination we can start performing the regressions. First we should check the normality of the data. We use two tests for checking normality of the data: the Shapiro-wilk w test for normal data (appendix, table 19) and the Kensel Density Estimation (appendix, figure 2). From these tests, we can conclude that the data is not normally distributed. Figure 2 shows that the dataset contains a large amount of zero's, namely 708. The p-value (0.000) from table 19 means that hypothesis H0: data is normally distributed needs to be rejected. As a result we cannot simple use a normal linear regression in our research design.

We continue the examination of the data with testing for the presence of heteroskedasticity in the dependent variables in this research. By means of residuals plot (appendix, figure 3 and Breusch-Pagan/Cook-Weisberg (appendix table 20) we are able to test for heteroskedasticity. Figure 3 shows that the variance is not uniformly spread around the red horizontal line. Table 20 shows that H0: equal variance needs to be rejected (p-value = 0.000). Therefore we can conclude that the data suffers from heteroskedasticity. As a result, robust standard errors are necessary in the regression analysis in order to make a correct adjustment for the heteroskedasticity. Figure 3 also gives some indication for having extremely large values (outliers) present in the dataset. As a result, we should first examine the outliers. Outliers are detected by using the rule of μ +- 2xσ (mean plus or minus two times the standard deviation). Because of this, 95% of the flights in the dataset will have a delay at arrival that is two standard deviations away from the average delay at arrival. From table 2, we find a mean of 10.77 for *delay_a_minutes* with a standard deviation of 37.96. Therefore a value is considered to be an outlier when it is lower than -65 and higher than 87. Removing outliers from the dataset leads to the loss of valuable data, because flights with a substantial delay at arrival are important for the outcomes of this research. Therefore it is valuable to keep the outliers. Winsorizing, as stated in footnote 15, is the appropriate method for the correction of outliers in this research. We find no negative outliers, however we do find 43 values that are higher than 87. This means that 43 flights have a delay of more than 87 minutes at arrival. These positive outliers are set to the value 87.

Finally, we check the variables on multicollinearity. To be short, there are no significant correlations between two variables that can lead to regression problems (appendix, table 21).

The analysis of this research starts with performing two-group mean comparison tests in order to get a first impression of the first results and the possible relation between the four independent variables *type_flight_carrier*, *Ryanair_other_low_cost_flight_carriers*, *Ryanair_other_carriers*, *Ryanair_KLM* and the dependent variable *delay_a_minutes(1)*.

Then we continue to expand the analysis by performing regressions. We start with a simple binary logistic regression of the main model. The binary logistic regression identifies whether or not an independent variable has a significant effect on the chance that a flight has delay at arrival. We continue with regressions of our four models based on the four hypotheses of this research. Each of the four models consists of the dependent variable *delay_a_minutes(2)*, a single independent variable and some control variables. Model 1 can be considered as the main model of this research. This model compares the on-time performance of low-cost carriers to high-cost/national flag carriers. Model 2, 3 and 4 are used to investigate the relative on-time performance of the low-cost carrier Ryanair. Model 2 evaluates the on-time performance of Ryanair and compares it to the on-time performance of all other carriers, so low-cost and high-cost carriers, of this research. Model 3 compares the on-time performance of Ryanair to the on-time performance of other low-cost carriers in this research. This model only uses the sample of flights from Ryanair and other low-cost flight carriers. Finally, model 4 relates the on-time performance of Ryanair to the on-time performance of the high cost/national flag carrier KLM. The sample of this model only consists of flights from Ryanair and KLM.

We identified the following models for this research:

Model 1 (main model) for binary logistic regression:
delay_a_minutes(0/1) = $\alpha + \beta_1$ x *type_flight_carrier* + β_2 x *time_sched_d* + β_3 x *type_airport_a* + β_4 x *date* + β_5 x *log_distance* + β_6 x *visibility_a1/2/3/4/5* + β_7 x *country_a1/2/3/4* + ε

Model 1 (main model):
delay_a_minutes(2) = $\alpha + \beta_1$ x *type_flight_carrier* + β_2 x *time_sched_d* + β_3x *type_airport_a* + β_4 x *date* + β_5 x *log_distance* + β_6 x *visibility_a1/2/3/4/5* + β_7 x *country_a1/2/3/4* + ε

Model 2:
delay_a_minutes(2) = $\alpha + \beta_1$ x *Ryanair_other_carriers* + β_2 x *time_sched_d* + β_3x *type_airport_a* + β_4 x *date* + β_5 x *log_distance* + β_6 x *visibility_a1/2/3/4/5* + β_7 x *country_a1/2/3/4* + ε

Model 3:
delay_a_minutes(2) = $\alpha + \beta_1$ x *Ryanair_other_low_cost_flight_carriers* + β_2 x *time_sched_d* + β_3 x *type_airport_a* + β_4 x *date* + β_5 x *log_distance* + β_6 x *visibility_a1/2/3/4/5* + β_7 x *country_a1/2/3/4* + ε

Model 4:
delay_a_minutes(2) = $\alpha + \beta_1$ x *Ryanair_KLM* + β_2 x *time_sched_d* + β_3 x *type_airport_a* + β_4 x *date* + β_5 x *log_distance* + β_6 x *visibility_a1/2/3/4/5* + β_7 x *country_a1/2/3/4* + ε

The dependent variable in the binary logistic regressions is *delay_a_minutes_0/1*. If a flight has a delay of less than 15 minutes, we say that the flight has no delay. On the other hand, a flight is delayed at arrival when the flights arrive 15 minutes or more behind the scheduled time of arrival. We will only perform a normal binary logistic regression and a normal binary logistic regression with robust standard errors of the main model with the independent variable *type_flight_carrier*. Control variables are added to the regression in order to see if and how these control variables change the effect of the independent variable in the binary logistic regression.

We perform more complex negative binomial and zero-inflated negative binomial regressions for all models. Previous paragraph 4.2 showed that the data in this research is non-normal. This non-normality is because the dataset contains a significant amount of 708 zero’s. As a result, a normal linear regression would not be the most appropriate regression for this research. A zero-inflated negative binomial model regression would be more suitable. In addition we will also perform some negative binomial model regressions, so that we can compare and evaluate the similarities and differences in results of the zero-inflated negative binomial model regression with the results of the negative binomial model regressions. Zero-inflated negative binomial and negative binomial regression are both count models and rely on almost the same assumptions. They both simulate the chance on a certain value. For example, given that the average delay at arrival (*delay_a_minutes(2)*) is 12 minutes, a delay of 12 minutes at arrival has the highest probability. Hereinafter, a delay at arrival of 11 or 13 minutes has the largest probability. However, the chance on a delay of zero minutes at arrival is relatively small, but since we have a lot of observations with the value zero, the chance on the value zero should be increased. The zero-inflated negative binomial model takes the large amounts of zero’s into account by separately modelling the chance on the value zero. In this research we use the logit function for modelling the chance on the value zero. Unlike the zero-inflated negative binomial model, the negative binomial model does not take the large amount of zero’s into account. Therefore it would be very interesting to see if there are differences between the two types of regressions. The dependent variable for the negative binomial and zero-inflated negative binomial model regressions is *delay_a_minutes(2)*. As a result of heteroskedasticity in the data, we will also perform negative binomial and zero-inflated negative binomial model regressions with robust standard errors. We take the zero-inflated negative binomial model regression with robust standard errors as base regression for the analyses, because we think this model produces the most reliable results.

5. Results

This section will describe the results of this research. The first analysis starts with the two-group mean comparison tests for the four independent variables. We use the variables without outliers for this analysis. The analysis continues with binary logistic regression results of the main model of this research. The last part of the analysis consists of negative binomial regressions and zero-inflated negative binomial regressions of all four models. Finally, there will be a discussion of the results. This discussion section relates our results to results of previous research and gives an evaluation of the hypotheses.

5.1 Mean group comparison tests

As previously mentioned, the analysis of this research starts with the two-group mean comparison tests. Table 3 shows the results for the two-group mean comparison test with *type_flight_carrier* as group variable. This two-group mean comparison test examines whether or not the mean delay in minutes at arrival of low-cost carriers differs from the mean delay in minutes at arrival of high-cost carriers. High-cost carriers and low-cost carriers are consecutively indicated with group 0 and 1 in table 3. The mean flight delay for high-cost carriers at arrival is 13.07 minutes, while the mean flight delay for low-cost carriers at arrival is 4.72 minutes. Due to this substantial difference in the mean of flight delay between the two groups, H0: diff = 0, needs to be rejected. As a result we can say that high-cost carriers have significant higher delay in minutes at arrival than low-cost carriers.

Table 3: Two group mean comparison test between high cost/national flag carriers (0) and low cost flight carriers (1)

Group	**Observations**	**Mean**	**St. Err.**
0	606	13.07	1.11
1	868	4.72	0.75
Diff = mean(0)-mean(1)		8.35	
Ha: diff < 0	**Ha: diff=0**		**Ha: diff > 0**
Pr(T<t) = 1.0000	**Pr(T<t) = 0.0000**		**Pr(T<t) = 0.0000**

The two-group mean comparison test in table 4 also has an interesting result. Ryanair denoted by group 1 has an average delay at arrival of 4.12 minutes, while the average delay for all other carriers, denoted by group 0 is 10.57 minutes. As a result, the average delay in minutes at arrival of Ryanair is significantly smaller than the average delay of all other carriers in this research (low- and high-cost carriers).

Table 4: Two group mean comparison test between low cost carriers without Ryanair or high cost/national flag carriers (0) and Ryanair (1)

Group	**Observations**	**Mean**	**St. Err.**
0	921	10.57	0.88
1	553	4.12	0.85
Diff = mean(0)-mean(1)		6.45	
Ha: diff < 0	**Ha: diff=0**		**Ha: diff > 0**
Pr(T<t) = 1.0000	**Pr(T<t) = 0.0000**		**Pr(T<t) = 0.0000**

The same type of analysis is done for the group with Ryanair and all other low-cost carriers. All low-cost flight carriers that are not Ryanair are indicated with group 0, while Ryanair flights are indicated with group 1. Table 5 gives the numbers for the average duration of flight delay at arrival for the two groups. All low-cost carriers without Ryanair have an average delay at arrival of 5.77 minutes. The average delay at arrival for Ryanair carriers is slightly lower, namely 4.12 minutes. Because of this small difference in means, H0: diff = 0, cannot be rejected. Therefore, there is no significant difference in the average flight delay at arrival of all low-cost carriers except Ryanair and the average flight delay at arrival for Ryanair.

Table 5: Two group mean comparison test between low cost flight carriers without Ryanair (0) and Ryanair (1)

Group	**Observations**	**Mean**	**St. Err.**
0	315	5.77	1.42
1	553	4.12	0.85
Diff = mean(0)-mean(1)		1.65	
Ha: diff < 0	**Ha: diff=0**		**Ha: diff > 0**
Pr(T<t) = 0.8548	**Pr(T<t) = 0.2904**		**Pr(T<t) = 0.1452**

The final two-group mean comparison test (table 6) is a test between the average delay at arrival of KLM (group 0) and the average delay at arrival of

Ryanair (group 1). The mean delay for KLM at arrival is 14.64 minutes, while the mean delay for Ryanair at arrival is only 4.12 minutes. The difference between the means of the two groups is more than ten minutes. The two-group mean comparison test shows that Ryanair has a significantly lower delay at arrival than KLM.

Table 6: Two group mean comparison test between KLM (0) and Ryanair (1)

Group	**Observations**	**Mean**	**St. Err.**
0	403	14.64	1.49
1	553	4.12	0.85
Diff = mean(0)-mean(1)		10.52	
Ha: diff < 0	**Ha: diff=0**		**Ha: diff > 0**
Pr(T<t) = 1.0000	**Pr(T<t) = 0.0000**		**Pr(T<t) = 0.0000**

Despite some interesting first results of the four two-group mean comparison tests, the results of these tests only provide a first impression about the possible effect of the different independent variables on on-time performance. These two-group mean comparison tests do not take the effect of possible control variables into account.

5.2 Binary logistic regression model 1

The binary logistic regression goes a step further than the previous two-group mean comparison tests. The binary logistic regression of the main model of this research provides us more information about variables that influence the chance whether or not a flight will have delay at arrival. Table 7 on the next page shows the results of the normal binary logistic regression and the binary logistic regression with robust standard errors for the main model with *delay_a_(0/1)* as dependent variable and *type_flight_carrier* as independent variable. The main results of the normal binary logistic regression and the binary logistic regression with robust standard errors are practically the same. The independent variable *type_ flight_carrier* has an insignificant negative effect. This means that probability on whether or not having a delay at arrival is lower for low-cost carriers than high-cost carriers, though the effect of *type_flight_carrier* is insignificant. This insignificant negative effect for *type_flight_carrier* is almost exactly the same for the binary logistic regression with robust standard errors. However, we do find highly significant effects of several control variables. *Time_sched_d*, *date* and *country dummies* all have highly significant effects. *Time_sched_d* has a significant positive effect. This means that flights with a later scheduled departure time have a higher probability on facing a flight delay. The variable *date* has a highly significant negative effect and therefore we can say that flights in December have significantly more chance on having a flight delay at arrival than flights in April. The country dummies *country_a2*, *country_a3* and *country_a4* all have a highly significant positive effect. This means that flights that arrived at a specific airport in Spain, Italy or Sweden have a higher probability on flight delay at arrival than flights that arrived at a specific airport in the Netherlands. *Time_sched_d*, *date* and country dummies have exactly the same coefficient sign, coefficient value and significance level in both the normal binary logistic regression and the binary logistic regression with robust standard errors. Finally, the variable *type_airport_a* has a significant positive effect. Therefore flight carriers that fly to main airports have a higher probability of facing delay at arrival than flight carriers that fly to secondary airports.

Table 7: Binary logistic regression of model 1 with *delay_a_(0/1)* as dependent variable. The independent variable is *type_flight_carrier*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

	Model 1		**Model 1 with robust standard errors**	
Variable	**Coefficient**	**Std. err.**	**Coefficient**	**Std. err.**
type_flight_carrier	-0.105	0.229	-0.105	0.231
time_sched_d	0.052***	0.017	0.052***	0.018
type_airport_a	0.488*	0.254	0.488*	0.251
date	-2.363***	0.193	-2.363***	0.192
log_distance	0.377	0.305	0.377	0.316
visibility_a1	Omitted		Omitted	
visibility_a2	0.352	0.391	0.352	0.404
visibility_a3	0.149	0.319	0.149	0.351
visibility_a4	0.040	0.235	0.040	0.235
country_a2	0.665***	0.241	0.665***	0.233
country_a3	0.700***	0.243	0.700***	0.249
country_a4	0.755***	0.261	0.755***	0.265
Observations	1168		1168	
Pseudo R^2	0.2100		0.2100	

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

5.3 Negative binomial model and zero-inflated negative binomial model regressions

We will now focus on the most comprehensive and most interesting results of this research. This part of analysis uses the more appropriate negative binomial model and zero-inflated negative binomial model regressions with and without robust standard errors. The dependent variable is *delay_a_minutes(2)*. The regressions provide more valuable information than the binary logistic regression in the previous part. We start with the regression results of the zero-inflated negative binomial model with robust standard errors for each of the four models. We consider this model to be the model that provides us with the most reliable results. After the results of this model we also perform the additional negative binomial model regressions with and without robust standard errors and the zero-inflated binomial model regression without robust standard errors to see how the results differ between the four different types of regressions.

Table 8 on the next page gives us the regression results of the zero-inflated negative binomial model with robust standard errors of the main model of this research. The independent variable in this main model is *type_flight_carrier*. *Type_flight_carrier* has a highly significant negative effect (coefficient value = -0.403). This means that low-cost carriers have less delay at arrival than high-cost carriers. However, the effect of *type_flight_carrier* becomes less significant when we add control variables to the regression. Eventually, *type_flight_carrier* has an insignificant positive effect (coefficient value = 0.140) on the delay at arrival. In other words, the type of flight carrier has a significant negative effect on the delay at arrival, though when we control for other variables, this significant negative effect changes into an insignificant positive effect. This positive coefficient sign of *type_flight_carrier* is not in line with our expectations, because we expected a better on-time performance of low-cost carriers. Other variables like *time_sched_d, type_airport_a* and *date* do have a significant effect on the delay at arrival of a flight. *Time_sched_d* has a significant positive effect (coefficient value = 0.032) on the delay at arrival of a flight carrier. Therefore flight carriers with a later scheduled flight time have more flight delay than flight carriers with earlier scheduled flight times. Also the type of airport is an important variable in this analysis.

Type_airport_a has a significant positive effect (coefficient value = 0.437). As a result, flight carriers that fly to main airports are expected to have more flight delay than flight carriers that fly to secondary airports. The variable *date* has a significant negative effect (coefficient value = -0.980). This means that flights in December have more delay than flights in April. Additionally, we find evidence for a significant positive effect of *visibility_a1*, *visibility_a3*, *country_a3* and *country_a4*. This means that very bad and poor visibility near the airport of arrival will cause more delay. However, the effect of the weather dummies provides little value, due to limited observations in this research. Finally, the *country_a3* and *country_a4* dummies have a significant positive effect. This means that on average, flights that arrive at airports in Italy and Sweden have more delay than flights that arrive at airports in The Netherlands.

Table 22 in the appendix shows the regression results of the negative binomial model with and without robust standard errors and the zero-inflated negative binomial model with and without robust standard errors. Most of the outcomes, like significance of variables and coefficient signs in the four different types of regressions are similar. However, there is quite some difference in coefficient values between the negative binomial model regressions and the zero-inflated negative binomial model regressions. For example, the coefficient value of *date* in the negative binomial model regressions is -1.603, while this variable has a coefficient value of -0.980 in the zero-inflated negative binomial model regressions. There is also some dissimilarity in standard errors between the regressions with and without robust standard errors. This indicates the presence of heteroskedasticity. The independent variable *type_flight_carrier* has an insignificant positive effect in each type of regression. The most significant variables are *time_sched_d*, *type_airport_a* and *date*. All these variables have a significant effect in each type of regression. Finally, we find some evidence for significant visibility and country dummies. However, these results are not very consistent.

Table 8: Zero-inflated negative binomial model regression with robust standard errors of model 1. The dependent variable is *delay_a_minutes(2)* and the independent variable is *type_flight_carrier*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Variable	**Coefficient (Std. err)**						
type_flight_carrier	-0.403*** (0.077)	-0.352*** (0.085)	0.064 (0.116)	0.101 (0.114)	0.106 (0.115)	0.131 (0.117)	0.140 (0.120)
time_sched_d		0.019** (0.009)	0.017* (0.009)	0.035*** (0.009)	0.035*** (0.009)	0.031*** (0.009)	0.032*** (0.009)
type_airport_a			0.633*** (0.133)	0.376*** (0.133)	0.379*** (0.132)	0.402*** (0.133)	0.437*** (0.136)
date				-0.960*** (0.080)	-0.961*** (0.081)	-0.960*** (0.091)	-0.980*** (0.089)
log_distance					0.030 (0.151)	0.119 (0.150)	0.103 (0.156)
visibility_a1						0.525*** (0.184)	0.729*** (0.211)
visibility_a2						0.090 (0.155)	0.276 (0.189)
visibility_a3						0.279 (0.191)	0.389** (0.197)
visibility_a4						-0.011 (0.108)	0.041 (0.114)
country_a2							0.199 (0.137)
country_a3							0.307** (0.142)
country_a4							0.277* (0.148)
Observations	1474	1221	1221	1221	1221	1173	1173

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Table 9 shows the regression results for model 2. This model compares the on-time performance of Ryanair to the on-time performance of all other carriers, so low-cost carriers and high-cost carriers. The results are largely identical to the results of the previous analysis that compared the delay at arrival of Ryanair to the delay at arrival of other low-cost carriers. First we find a significant negative effect of *Ryanair_other_carriers* (coefficient value = -0.441) in the regression without control variables. However, the negative effect of *Ryanair_other_carriers* completely disappears, when *time_sched_d*, *type_airport_a* and *date* are also in the regression. Finally, the variable *Ryanair_other_carriers* has a significant positive effect (coefficient value = 0.391). This indicates that Ryanair has more delay than all other low-cost and high-cost flight carriers in our sample. Again we find high significance of *time_sched_d*, *type_airport_a* and *date*. In these regressions we find some evidence for significant weather and country effects too.

Table 23 in the appendix shows that there is some difference between the results of the different types of regressions. *Ryanair_other_carriers* has a greater effect in the negative binomial model regressions (coefficient value = 0.664) than in the zero-inflated negative biniomial regressions (coefficient value = 0.391). Moreover, the effect of *Ryanair_other_carriers* is more significant (significant at the 0.05 level) in the negative binomial model regressions than in the zero-inflated negative binomial model regressions (significant at the 0.10 level). Here, we find some differences in standard errors between different regressions of the same type too.

Table 9: Zero-inflated negative binomial model regression with robust standard errors of model 2. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_other_carriers*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Variable				**Coefficient (Std. err)**			
Ryanair_other_carriers	-0.441*** (0.088)	-0.564*** (0.090)	-0.326* (0.190)	0.285 (0.197)	0.331 (0.214)	0.361* (0.217)	0.391* (0.230)
time_sched_d		0.018** (0.009)	0.017** (0.009)	0.035*** (0.009)	0.036*** (0.010)	0.033*** (0.009)	0.034*** (0.009)
type_airport_a			0.281 (0.202)	0.546*** (0.202)	0.581*** (0.210)	0.609*** (0.209)	0.672*** (0.223)
date				-0.995*** (0.084)	-1.008*** (0.088)	-1.010*** (0.098)	-1.030*** (0.096)
log_distance					0.100 (0.162)	0.194 (0.156)	0.198 (0.165)
visibility_a1						0.545*** (0.192)	0.736*** (0.219)
visibility_a2						0.103 (0.152)	0.278 (0.187)
visibility_a3						0.273 (0.196)	0.375* (0.202)
visibility_a4						-0.016 (0.106)	0.031 (0.113)
country_a2							0.162 (0.141)
country_a3							0.314** (0.142)
country_a4							0.250 (0.148)
Observations	1474	1221	1221	1221	1221	1173	1173

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Table 10 show the results of the zero-inflated model regression with robust standard errors for model 3 with *delay_a_minutes(2)* as dependent variable and *Ryanair_other_low_cost_carriers* as independent variable. The variable *Ryanair_other_low_cost_carriers* has a significant negative effect (coefficient value = -0.263), when no control variables are used. This implies that Ryanair has less delay at arrival than other low-cost carriers. Surprisingly, this effect of *Ryanair_other_low_cost_carriers* becomes greater (coefficient value = -0.766), when we add the control variable *time_sched_d*. The effect of the independent variable slightly weakens when we add the second control variable *type_airport_a*. However, the effect of *Ryanair_other_low_cost_carriers* completely changes when we also check for the variable date. The variable *Ryanair_other_low_cost_carriers* turns out to be insignificant and positive (coefficient value = 0.395) and this effect remains almost the same (insignificant coefficient value = 0.325), when all control variables are in the regression. Therefore on average, Ryanair has even more delay than other low-cost carriers. Yet, the effect is insignificant. The last column of table 10 again demonstrates that time of scheduled departure, type of airport and *date* have a significant effect on the delay at arrival. To be more specific, *time_sched_d* has a significant positive effect, *type_airport_a* also has a significant positive effect and *date* has a significant negative effect on flight delay. *Visibility_a1*, *country_a3* and *country_a4* also have a significant effect.

We can see from table 24 that the results of model 3 of the negative binomial model regressions differ from the results of the zero-inflated negative binomial regressions. The most important difference between the two types of regressions is that the variable *Ryanair_other_low_cost_carriers* has a significant positive effect in the negative binomial model regressions (coefficient value = 0.959), while the same variable is insignificant in the zero-inflated negative binomial model regressions. The results of the zero-inflated negative binomial model regressions are more reliable, because the negative binomial model regressions do not take the effect of the large amount of zero's into account. We also find that other coefficient values of the two types of regressions differ substantially. To be more specific, we find a greater effect of most variables in the negative binomial model regressions than in the zero-inflated negative binomial model regressions.

Table 10: Zero-inflated negative binomial model regression with robust standard errors of model 3. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_other_low_cost_flight_carriers*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category

Variable	**Coefficient (Std. err)**						
Ryanair_other_low_cost_flight_carriers	-0.263** (0.121)	-0.766*** (0.141)	-0.535** (0.219)	0.395 (0.242)	0.318 (0.292)	0.287 (0.291)	0.325 (0.295)
time_sched_d		0.021 (0.134)	0.020 (0.014)	0.032** (0.015)	0.030* (0.015)	0.024* (0.014)	0.028** (0.014)
type_airport_a			0.272 (0.201)	0.563*** (0.203)	0.516** (0.227)	0.523** (0.220)	0.665*** (0.242)
date				-1.098*** (0.143)	-1.074*** (0.151)	-1.004*** (0.170)	-1.056*** (0.167)
log_distance					-0.127 (0.281)	-0.077 (0.258)	-0.258 (0.268)
visibility_a1						0.893*** (0.152)	1.069*** (0.211)
visibility_a2						0.251 (0.194)	0.411 (0.256)
visibility_a3						0.340 (0.286)	0.382 (0.289)
visibility_a4						-0.142 (0.165)	-0.142 (0.161)
country_a2							0.095 (0.221)
country_a3							0.371* (0.219)
country_a4							0.522** (0.256)
Observations	868	655	655	655	655	616	616

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Finally, table 11 on the next page shows the results of the regression of model 4. This model compares the on-time performance of the low-cost carrier Ryanair and the high-cost carrier KLM. The initial regression results of the zero-inflated binomial model with robust standard errors in table 11 indicate that there is a significant negative effect (coefficient value = -0.629) of the independent variable *Ryanair_KLM*. This means that Ryanair has significantly less delay at arrival than KLM. However, this effect changes after controlling for other variables. Eventually, *Ryanair_KLM* has a significant positive effect, which means that Ryanair has significantly more delay than KLM (coefficient value = 0.522). Also, in this regression there is a significant effect of scheduled time of departure (coefficient value = 0.033), type of airport (coefficient value = 0.713) and *date* (coefficient value = -1.115). Additionally, we also find some evidence for significant weather and country effects.

Additional negative binomial and zero-inflated negative binomial model regressions (appendix, table 25) basically show the same results. However, there are some differences between the two types of regressions. The variable *Ryanair_KLM* has a greater effect in the negative binomial model regressions than in the zero-inflated negative binomial model regressions. Most other variables also have a greater effect in the negative binomial model regressions than in the zero-inflated negative binomial model regressions. Finally, the two types of regressions have some dissimilarity in the significance of the visibility and country dummies.

Table 11: Zero-inflated negative binomial model regression with robust standard errors of model 4. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_KLM*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Variable	**Coefficient (Std. err)**						
Ryanair_KLM	-0.629*** (0.096)	-0.634*** (0.096)	-0.396** (0.193)	0.449** (0.218)	0.452* (0.242)	0.483** (0.243)	0.522** (0.260)
time_sched_d		0.020** (0.010)	0.020** (0.010)	0.035*** (0.011)	0.035*** (0.011)	0.032*** (0.010)	0.033*** (0.010)
type_airport_a			0.280 (0.202)	0.575*** (0.203)	0.577*** (0.224)	0.609*** (0.216)	0.713*** (0.234)
date				-1.131*** (0.115)	-1.131*** (0.116)	-1.133*** (0.128)	-1.115*** (0.125)
log_distance					0.007 (0.286)	0.101 (0.273)	0.133 (0.257)
visibility_a1						0.462** (0.220)	0.619** (0.247)
visibility_a2						0.046 (0.182)	0.190 (0.217)
visibility_a3						0.305 (0.220)	0.392* (0.228)
visibility_a4						0.040 (0.117)	0.073 (0.123)
country_a2							0.095 (0.163)
country_a3							0.295* (0.158)
country_a4							0.219 (0.160)
Observations	956	956	956	956	956	916	916

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

5.4 Discussion of Results

In this part we will relate the results of this research to results of previous research. Consecutively, there will be an evaluation of the four hypotheses of this research. The evaluation of the hypotheses is based on the results of the zero-inflated negative binomial model regressions with robust standard errors from paragraph 5.3.

The final results of the main model (model 1) indicate that scheduling decisions are important. Flights that have later scheduled flight times have significantly more flight delay than flights with earlier scheduled flight times. This result is in line with the findings by Mazzeo (2003). Mazzeo (2003) also finds evidence for the importance of scheduling decisions. To be more specific, Mazzeo (2003) finds that the effect of a flight scheduled arrival time is quite large: the average flight arriving at 8 pm is 9 minutes more behind scheduled arrival than flights with a scheduled arriving time of 8 am. We also find a significant effect of the type of airport. Flights that arrive at a main airport have significantly more delay than flights that arrive at a secondary airport. Although no other research has examined the type of airport as independent variable exactly, there is still some research that uses similar variables. For example Mazzeo (2003) finds that congestion has a significant positive effect on flight delay. This means that more congestion leads to more flight delay. It is often argued that main and secondary airports differ in the amount of air traffic they process. Main airports have substantially more traffic and therefore congestion is more of a problem at main airports. Our outcome that flights have more delay at arrival at main airports is in some way similar to the significant congestion effect of Mazzeo (2003). Another variable that has a highly significant effect on the delay at arrival of a flight is *date*. We find that flights in December 2010 had more delay than flights in April 2011. This can be the result of seasonal effects. In December, a lot of holidays take place and as a result there can be more air traffic in this month. December also has more exposure to some extreme weather conditions. In contrast to December, April is a relatively quiet month and does not often have to deal with extreme weather conditions. Rupp et al. (2001) also find evidence for such seasonal effects. They find that winter flights experience recurrent and

longer flight delays and that summer flights also have more delay. Furthermore we also find some significant influence of the visibility dummy 'very bad visibility'. However, these visibility dummies do not provide a value finding due to limited observations. Yet, there is still a large possibility that weather conditions do matter in explaining flight delays. Mazzeo (2003) also mentioned the importance of weather conditions in clarifying the mystery about flight delays as he finds highly significant weather variables, like thunder, snow, rain or fog. The last control variable that has a significant influence on delay at arrival is the country Italy. Flying to Italy will lead to more delay compared with flights that go to The Netherlands. This difference in on-time performance can be caused by differences in the quality of the airline infrastructure of the two countries.
The first hypothesis that is related to the main model is:

H1: there is no difference between the on-time performance of low-cost carriers and the on-time performance of high-cost carriers.

We find an insignificant positive coefficient of the variable type of flight carrier after controlling for several other important variables. As a result, there is no difference between the on-time performance of low-cost carriers and the on-time performance of high-cost carriers. Therefore, we fail to reject hypothesis H1.

We use the models 2, 3 and 4 for the evaluation of hypotheses H2, H3 and H4. These models are practically the same as model 1. All the models use the same types of regressions and the same control variables. The only difference between the models is the use of the independent variable. The regression results of model 2, 3 and 4 show that time of scheduled departure, type of airport at arrival and *date* in each regression has a significant effect on the delay at arrival. Furthermore we find inconsistencies in the significance of some visibility and country dummies. The hypothesis based on model 2 is:

H2: there is no difference between the on-time performance of Ryanair and the on-time performance of other low-cost carriers and high-cost carriers.

The regression results of model 2 show a significant positive effect of the variable Ryanair versus all other carriers. Therefore, there is a difference in on-time performance between Ryanair and other low-cost and high-cost carriers. To be more specific, the on-time performance of Ryanair is worse than the on-time performance of the sample with all other flight carriers in this research. As a result, we find support to reject hypothesis H2.

Model 3 compares the difference in on-time performance between Ryanair and other low-cost carriers. The third hypothesis, related to model 3, is:

H3: there is no difference between the on-time performance of Ryanair and the on-time performance of other low-cost carriers.

For model 3 we find that the variable of Ryanair versus other low-cost carriers has an insignificant effect. This means that there is no evidence for a significant difference between the on-time performance of Ryanair and the on-time performance of other low-cost carriers in this research. Thus, we fail to reject hypothesis H3.

Model 4 compares the on-time performance of Europe's largest low-cost carriers to the on-time performance of the well-known national flag/high-cost carrier KLM. The final hypothesis is related to model 4:

H4: there is no difference between the on-time performance of Ryanair and the on-time performance of KLM.

Final regression results show a significant positive effect of the variable Ryanair versus KLM. This means that the on-time performance of Ryanair is significantly worse than the on-time performance of KLM. Therefore, we find support to reject hypothesis H4.

6. Conclusion

The last chapter of this research starts with a general conclusion. This general conclusion answers the research question we addressed in the introduction. After the general conclusion some strategy implications for flight carriers will be given. The final part of the conclusion will address some limitations of this research and possibilities for future research.

6.1 General Conclusion

In the introduction we defined the following research question:

Is there a difference in on-time performance between low-cost carriers and high-cost carriers after controlling for some important factors?

By means of appropriate zero-inflated negative binomial model regressions with robust standard errors, we find that low-cost flight carriers have significantly better on-time performance than high-cost cost carriers, when we do not control for other variables. However, we find no significant difference in on-time performance between low-and high-cost carriers when we do control for scheduling decisions, different types of airports, seasonal effects, weather effects and country effects. This means that high-cost carriers are able to achieve the same on-time performance as low-cost carriers, if they have the same scheduled departure times, fly to the same type of airports and have to deal with the same seasonal effects as low-cost carriers. We also examined the relative performance of Europe's largest low-cost flight carrier Ryanair to other low- and high-cost flight carriers. Initially, the different regressions show that Ryanair has a significantly better on-time performance than other low-cost and high-cost flight carriers. However, this result completely changes, when we control for other variables. We find that the on-time performance of Ryanair is significantly worse than the on-time performance of the sample with other low-cost and high cost carriers. However, we find no significant difference in on-time performance, if we compare the on-time performance of Ryanair to the on-time performance of only low-cost carriers. Finally, our results indicate that there is a significant difference between the on-time performance of Ryanair and the on-time performance of KLM. To be more specific, the on-time performance of Ryanair is significantly worse than the on-time performance of KLM, after controlling for scheduling decisions, different types of airports, seasonal effects, weather effects and country effects. Therefore, the initial better performance of Ryanair compared to other flight carriers is because Ryanair primarily has early scheduled departure times, generally flies to secondary airports and has less interference from seasonal effects.

6.2 Strategy Implications

While many airlines argue that the causes of flight delays are outside their control, this research finds some counter facts. The most important implication of this research, based on the general conclusion, is that high-cost carriers theoretically can achieve the same or even a better on-time performance than low-cost carriers. Better on-time performance of low-cost carriers is caused by the differences in scheduling decisions, type of airports, seasonal effects, weather effects and country effects. Some causes, like seasonal and weather effects, are not really controllable by firms. Though, other factors like scheduling flight times and type of airport are manageable for airliners. On-time performance is a very important performance measure for both passengers and flight carriers. Previous research found that passengers are more likely to switch to another airliner if they experienced flight delays with a specific airliner in the past. Therefore, especially high-cost airliners should really take into account that the effects of their decisions about flight schedules and type of airports can really affect their on-time performance and indirectly also the financial performance. High-cost carriers with poor on-time performance could lose passengers to low-cost carriers with a better on-time performance. Flying fewer passengers will have a negative impact on revenues and other financial performance measures. Of course, there is a huge difference in strategy, goals and markets of low-and high cost carriers. Low-cost carriers focus primarily on frequent continental low-cost point-to-point flights. They more or less offer a mass product, by not focussing on the individual passenger with personal specific demands, but rather on a large group of passengers with common requirements. These common requirements consist of passengers that want frequent and low-cost flights. In contrast to low-cost carriers, most high-cost carriers perform both continental and intercontinental flights. High-cost carriers ascribe a more central role for the individual. They want to meet the demands of the individual passenger. This individual passenger wants to have more options and flexibility regarding his flights. High-cost carriers could use their high quality networks of direct continental and intercontinental air connections to satisfy the needs of this specific type of passenger. The risk of competition from low-cost carriers for high-cost carriers is the highest on continental routes. Low-cost carriers can more easily compete on continental flights than on intercontinental flights. This is because the difference in service level between low- and high-cost carriers on continental flights is not that spectacular and most passengers on continental flights have little demands. As a result, price, frequency of flights on a specific route and on-time performance can be crucial factors for passengers to either choose for a low- or a high cost carrier. Therefore, high-cost carriers should really take the effects of these crucial factors on their competitive resistance against low-cost carriers into account.

6.3 Limitations and Future research

Although this research made a good first attempt to investigate the relationship between the type of carrier and on-time performance, there are also some limitations concerning this research. The number of observations (flights) in the regressions was initially quite large. However, for some model regressions, the total number of observations has dropped significantly. This decrease in the total number of observations in some models is primarily due to sample restrictions. Also, the total number of observations in December 2010 was significantly lower than the total number of observations in April 2011. Therefore, seasonal and weather effects have less explanatory power. Moreover, this research has used some very important control variables, though the total number of control variables is somewhat limited. Finally, it would be valuable to look at more flights from different flight carriers, flight routes, countries and from different years.

This research and earlier studies have tried to identify factors that are related to the on-time performance of flight carriers. While we examined the difference in on-time performance between low-cost carriers and high-cost carriers, others came up with other interesting research subjects. For example, Suziki (2000) related market share of flight carriers to on-time performance. Both Rupp et al. (2001) and Mazzeo (2003) examined the effect of competition on on-time performance of flight carriers. Additionally, a more recent study by Prince and Simon (2009) investigated the relationship between multimarket contact and on-time performance in the airline industry. All of this research found significant results. These results indicate that on-time performance of flight carriers is probably related to a lot of variables that are currently not identified yet. Future research can contribute significantly, by expanding research on on-time performance based on existing studies. Probably, the most comprehensive empirical research on on-time performance is from Rupp et al. (2001). They used a wide range of (control) variables in their research on the relation between competition and on-time performance. Future research should combine methodologies of current research and should also use a wide range of control variables in order to control for specific effects. It would also be interesting to see how for example the use of panel data affects the existing results on on-time performance. To summarize, current research about on-time statistics has generated some valuable first results, however this field of study needs a lot more research in order to solve the ambiguities of on-time performance in the airline industry.

7. References

Bamberger, G.E., Carlton, D.W. and Neumann, L.R., An Emperical Investigation of the Competitive Effects of Domestic Airline Alliances, National Bureau of Economic Research, 2001

Borenstein, S. and Rose, N.L., Competition and Price Dispersion in the U.S. Airline Industry, National Bureau of Economic Research, 1991

Borenstein, S and Rose, N.L., How Airline Market Work...Or Do They? Regulatory Reform In The Airline Industry, National Bureau of Economic Research, 2007

Bowen, B.D. and Headley, D.E, The Airline Quality Rating 2001, The Journal of Aviation, 2001

Burghouwt, G. and Huys, M., Deregulation and the Consequences for Airport Planning in Europe, disP, 2003, Vol. 153, No. 3, pp. 37-45

Burton, J. and Hanlon, P., Airline alliances: cooperating to compete?, Journal of Air Transport Management 1995, pp. 209-227

Franke, M., Competition between network carriers andlow-cost carriers— retreat battle or breakthrough to a new level of efficiency?, Journal of Air Transport Management, 2004, pp.15-21

Hurdle, G.J., Johnson, R.L., Joskow, A.S., Werden, G.J. and Williams, M.A., Concentration, Potential Entry, and Performance in the Airline Industry, The Journal of Industrial Economics, 1989, Vol. 38, pp. 119-139

Ito, H., and Lee, D., Low Cost Carrier Growth in the US Airline Industry: Past, Present, and Future, Working Paper, University of Brown, Department of Economics, 2003

Keeler, T.E., Airline Regulation and Market Performance, The Bell Journal of Economics and Management Science, 1972, Vol.3, No. 2, pp. 399-424

Oum, T.H., Zhang, A. and Fu, X., Air transport liberalization and its impact on airline competition and air passenger traffic, Transportation Journal, 2010, Vol. 49, No. 4, pp. 24-41

Mazzeo, M.J., Competition and Service Quality in the U.S. Airline Industry, Review of Industrial Organization, 2003, 22: pp. 275-296

Prince, J.T. and Simon D.H., Multimarket Contact and Service Quality: Evidence from On-Time Performance in the U.S. Airline Industry, Academy of Management Journal, 2009, Vol. 52, No. 2, pp. 336-353

Rhoades, D.L. and Lush, H., A typology of strategic alliances in the airline industry: Propositions for stability and duration, Journal of Air Transport Management, 1997, Vol. 3, No. 3, pp. 109-114

Rupp, N.G., Owens, D. and Plumly, L.W., Does Competition Influence Airline On-Time Performance?, Working Paper, Universtity of East Carolina, Department of Economics, 2001

Suziki, Y., The relationship between on-time performance and airline market share: a new approach,Transportation Research Part E, 2000, pp. 139-154

8. Appendix

Figure 1: Domestic market share based on domestic revenue passenger-miles of Southwest and other low-cost carriers in the U.S. in the period of 1984-2005.

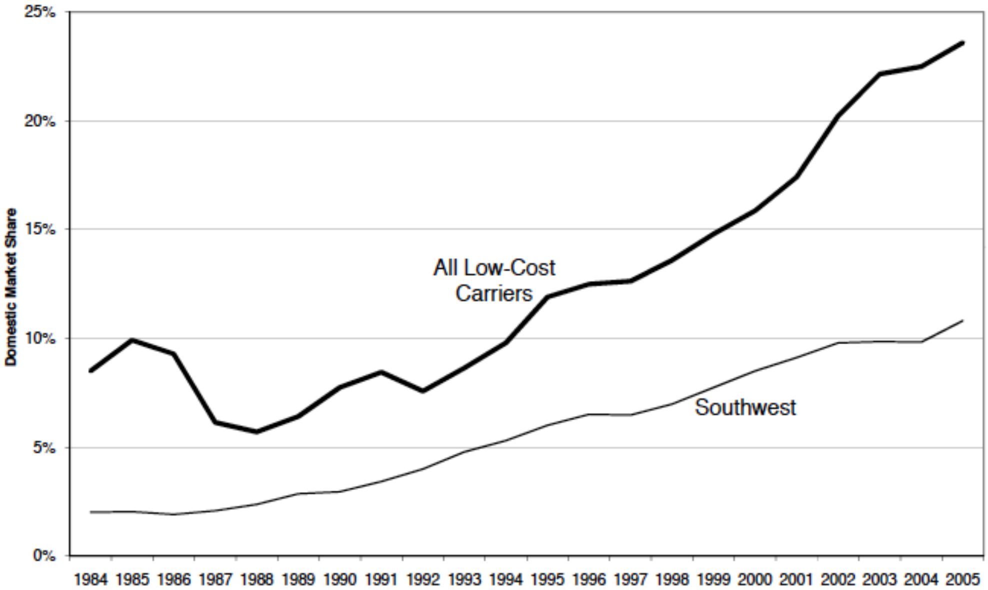

Table 12: Classification of causes of airports delay by the Bureau of Transportation Statistics (BTS) in the U.S. There are five categories: air carrier, extreme weather, National Aviation System (NAS), late-arriving aircraft and security.

Air carrier	The cause of the cancellation or delay was due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
Extreme weather	Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as a tornado, blizzard or hurricane.
National Aviation System (NAS)	Delays and cancellations attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume and air traffic control.
Late-arriving aircraft	A previous flight with same aircraft arrived late, causing the present flight to depart late.
Security	Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

Table 13: The distribution of causes of delay in percentages

Cause of delay	**Year**							
	2003 (Jun-Dec)	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Air Carrier Delay	26,3%	25,8%	28,0%	27,8%	28,5%	27,8%	28,0%	30,4%
Aircraft Arriving Late	30,9%	33,6%	34,2%	37,0%	37,7%	36,6%	36,2%	39,4%
Security Delay	0,3%	0,3%	0,2%	0,3%	0,2%	0,1%	0,1%	0,2%
National Aviation System Delay	36,5%	33,5%	31,4%	29,4%	27,9%	30,2%	30,6%	25,7%
Extreme Weather	6,1%	6,9%	6,2%	5,6%	5,7%	5,4%	5,0%	4,4%

Table 14: Flight routes

Flight route	Number of flights for given route	Distance of flight route (kilometres)	Flight route	Number of flights for given route
AMS-ARN	68	1154	ARN-AMS	71
AMS-BCN	181	1242	BCN-AMS	73
AMS-FCO	118	1298	FCO-AMS	68
AMS-GRO	20	1168		
ARN-BCN	39	2318	BCN-ARN	39
ARN-FCO	32	2024	FCO-ARN	41
BCN-CIA	56	877	CIA-BCN	60
BCN-FCO	44	848	FCO-BCN	86
CIA-EIN	47	1206	EIN-CIA	46
CIA-GRO	13	814	GRO-CIA	13
CIA-NYO	21	1915	NYO-CIA	21
EIN-GRO	30	1082	GRO-EIN	30
EIN-NYO	28	1095	NYO-EIN	28
EIN-REU	21	1191	REU-EIN	21
GRO-MST	9	1030	MST-GRO	9
GRO-NYO	38	2122	NYO-GRO	36
MST-REU	13	1144	REU-MST	13
RTM-BCN	19	1200		
RTM-FCO	22	1275		
			Total flight routes: 35	**Total flights: 1474**

Table 15: Country statistics

Country	Number of flights arrived at an airport in the specific country
The Netherlands	360
Spain	527
Italy	352
Sweden	235

Table 16: Type of airport statistics

Type of airport	Number of arrivals at specific type of airport	Number of delayed arrivals at specific type of airport	Percentage delayed arrivals at specific type of airport
Main	961	533	55.46%
Secondary	513	233	45.42%
Total	1474	766	51.97%

Table 17: Flight carrier descriptive statistics; number of flights per flight carrier, the mean delay at arrival in minutes and the standard deviation in minutes

Flight carrier	**Type of carrier (low vs. high cost)**	**Observations (number of flights)**	**Mean delay arrival in minutes[17]/[18]/[19]**			**Standard deviation in minutes ([16/17/18])**		
Alitalia	High	83	6.43	6.29	10.37	21.56	20.96	17.70
EasyJet	Low	74	20.39	14.19	17.91	53.68	29.75	26.65
KLM	High	403	18.20	14.64	18.55	41.85	29.96	26.68
Norwegian Air	Low	59	-5.07	-5.07	3.47	14.24	14.24	8.08
Ryanair	Low	553	6.17	4.12	8.48	33.82	20.09	16.98
SAS	High	120	12.75	12.51	15.15	21.73	20.71	17.81
Transavia	Low	81	14.56	6.27	10.27	62.66	25.04	22.72
Vueling	Low	101	6.77	5.52	11.64	30.98	24.44	19.31
		Total: 1474						

Table 18: Airport statistics; the number of passengers a specific airport has and the type of the specific airport. The type of airport has two different categories: main and secondary.

Airports	**Number of passengers per year (year)**	**Type of airport**
The Netherlands		
Amsterdam Schiphol (AMS)	45211749 (2010)	Main
Rotterdam the Hague (RTM)	1000858 (2010)	Secondary
Eindhoven Airport (EIN)	2142832 (2010)	Secondary
Maastricht Aachen Airport (MST)	282000 (2008)	Secondary
Spain		
Barcelona Airport (BCN)	29209595 (2010)	Main
Barcelona Girona Airport (GRO)	4863785 (2010)	Secondary
Reus Airport (REU)	1421341 (2010)	Secondary
Italy		
Rome Fiumicino (FCO)	36228490 (2010)	Main
Rome Ciampino Airport (CIA)	4680765 (2010)	Secondary
Sweden		
Stockholm-Arlanda Airport (ARN)	16962416 (2010)	Main
Stockholm Skavsta (NYO)	2513046 (2010)	Secondary

[17] The mean of delay in minutes of a flight at arrival

[18] The mean of delay in minutes of a flight at arrival (*delay_ a_minutes(1)*). Outliers are winsorized by means of: μ +- 2x σ. μ and σ are derived from the mean and the standard deviation of the delay in minutes at arrival (see footnote [14])

[19] The mean of delay in minutes of a flight at arrival without outliers (see footnote [15]) and without negative values (*delay_a_minutes(2)*), because negative can not be used in the zero-inflated model negative binomial model. Negative values are set to zero.

Table 19: Shapiro-wilk w test for normal data

Variable	Obs	w	v	z	Prob > z
r	1151	0.633	263.229	13.880	0.000

Figure 2: Kensel Density Estimation of delay_a_minutes

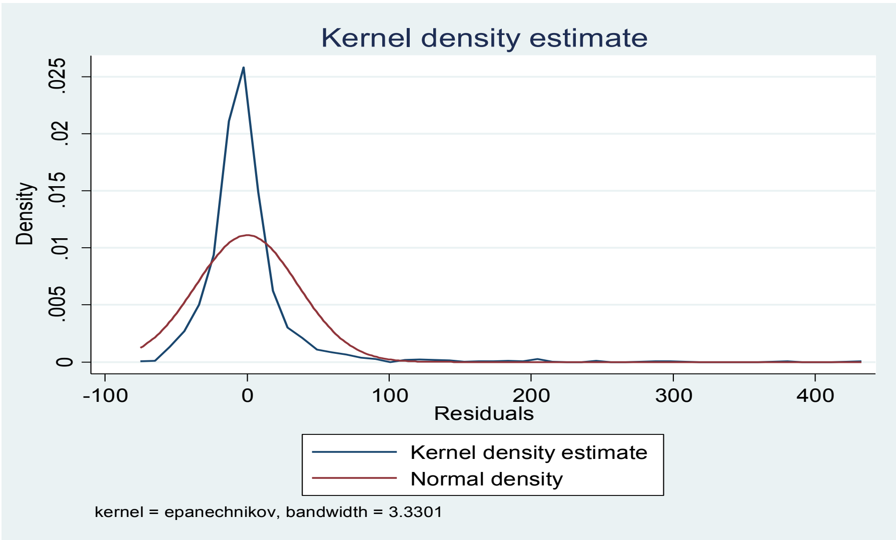

Table 20: Breusch-Pagan/Cook-Weisberg test for heteroskedasticity

H0: constant variance
Variables: fitted values of delay_a_minutes
Chi2(1) = 493.85
Prob>Chi2 = 0.000

Figure 3: Residuals plot of delay_a_minutes

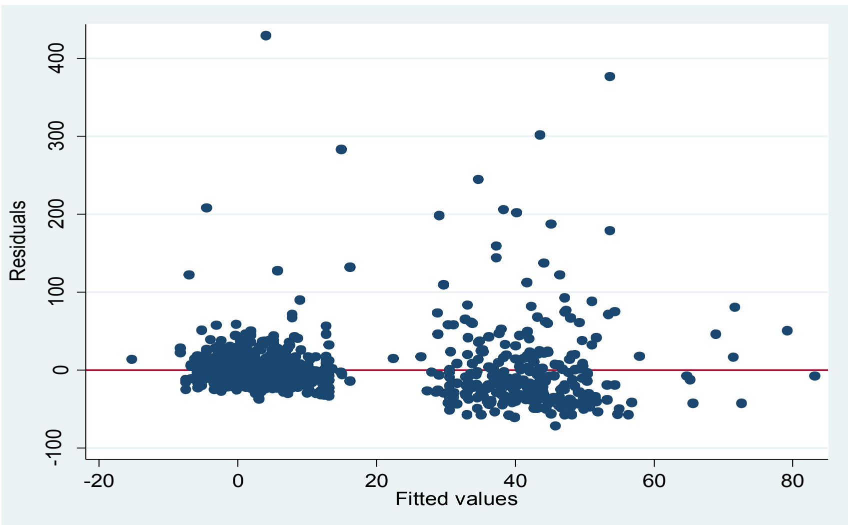

Table 21: Correlations

	Date	Time_sched_d	Delay_a_minutes_0/1	Delay_a_minutes (1)	Delay_a_minutes (2)	Type_airport_a	Log_distance	Visibility_a1	Visibility_a2	Visibility_a3	Visibility_a4	Visibility_a5	Country_a1	Country_a2	Country_a3	Country_a4
Date	1,0000															
Time_sched_d	0,0409	1,0000														
Delay_a_minutes_0/1	-0,4347	0,0865	1,0000													
Delay_a_minutes (1)	-0,4916	0,0902	0,7497	1,0000												
Delay_a_minutes (2)	-0,4916	0,0902	0,7497	1,0000	1,0000											
Type_airport_a	-0,2566	0,0582	0,2584	0,2467	-0,2467	1,0000										
Log_distance	0,0573	-0,1183	0,0377	-0,0131	-0,0131	-0,0006	1,0000									
Visibility_a1	-0,1149	-0,0652	0,0975	0,1331	0,1331	0,0648	0,0025	1,0000								
Visibility_a2	-0,2246	-0,0311	0,0990	0,1290	0,1290	0,1243	-0,0091	-0,0188	1,0000							
Visibility_a3	-0,2743	-0,0652	0,1099	0,1076	0,1076	-0,0296	0,0107	-0,0246	-0,0634	1,0000						
Visibility_a4	-0,2011	-0,0762	0,0531	0,0589	0,0589	0,0171	-0,1413	-0,0387	-0,0998	-0,1306	1,0000					
Visibility_a5	0,4481	0,1249	-0,1706	-0,1944	-0,1944	-0,0657	0,1134	-0,1301	-0,3359	-0,4393	-0,6916	1,0000				
Country_a1	-0,2882	-0,0910	0,0723	0,0978	0,0978	0,0349	-0,0685	0,1450	0,2797	0,2656	0,1024	-0,3929	1,0000			
Country_a1	0,1795	0,1142	-0,0483	-0,0321	-0,0321	0,1661	-0,1053	-0,0561	-0,0999	-0,1658	-0,0890	0,2253	-0,3871	1,0000		
Country_a1	0,0192	-0,0235	-0,0057	-0,0247	-0,0247	-0,1580	-0,2111	-0,0483	-0,1086	-0,1251	0,0480	0,0919	-0,3332	-0,3725	1,0000	
Country_a1	0,0887	-0,0065	-0,0175	-0,0437	-0,0437	-0,0596	0,4237	-0,0424	-0,0749	0,0342	-0,0613	0,0726	-0,2924	-0,3268	-0,2813	1,0000

Table 22: Negative binomial model regression (NBM), negative binomial model regression with robust standard errors, zero-inflated negative binomial model regression (ZINB) and zero-inflated negative binomial model regression with robust standard errors of model 1. The dependent variable is *delay_a_minutes(2)* and the independent variable is *type_flight_carrier*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Model 1	**NBM**		**NBM robust standard errors**		**ZINB**		**ZINB robust standard errors**	
Variable	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**
type_flight_carrier	0.051	0.198	0.051	0.156	0.140	0.122	0.140	0.120
time_sched_d	0.055***	0.013	0.055***	0.012	0.032***	0.008	0.032***	0.009
type_airport_a	0.367*	0.212	0.367**	0.173	0.437***	0.136	0.437***	0.136
date	-1.603***	0.166	-1.603***	0.112	-0.980***	0.093	-0.980***	0.089
log_distance	-0.087	0.236	-0.087	0.197	0.103	0.156	0.103	0.156
visibility_a1	1.324	0.928	1.324***	0.234	0.729*	0.433	0.729***	0.211
visibility_a2	0.614*	0.364	0.614***	0.224	0.276	0.200	0.276	0.189
visibility_a3	0.523*	0.273	0.523**	0.263	0.389**	0.163	0.389**	0.197
visibility_a4	0.240	0.184	0.240*	0.140	0.041	0.109	0.041	0.114
country_a2	0.550***	0.176	0.550***	0.165	0.199	0.122	0.199	0.137
country_a3	0.628***	0.176	0.628***	0.169	0.307**	0.123	0.307**	0.142
country_a4	0.839***	0.214	0.839***	0.192	0.277**	0.136	0.277*	0.148
Observations	1173				1173			
Pseudo R^2	0.027							

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Table 23: Negative binomial model regression (NBM), negative binomial model regression with robust standard errors, zero-inflated negative binomial model regression (ZINB) and zero-inflated negative binomial model regression with robust standard errors of model 2. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_other_carriers*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Model 2	**NBM**		**NBM robust standard errors**		**ZINB**		**ZINB robust standard errors**	
Variable	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**
Ryanair_other_carriers	0.664**	0.319	0.664**	0.270	0.391*	0.199	0.391*	0.230
time_sched_d	0.060***	0.013	0.060***	0.012	0.034***	0.008	0.034***	0.009
type_airport_a	0.924***	0.322	0.924***	0.266	0.672***	0.201	0.672***	0.223
date	-1.678***	0.168	-1.678***	0.119	-1.030***	0.096	-1.030***	0.096
log_distance	0.112	0.248	0.112	0.208	0.198	0.165	0.198	0.165
visibility_a1	1.314	0.924	1.314***	0.236	0.736*	0.432	0.736***	0.219
visibility_a2	0.590	0.363	0.590***	0.223	0.278	0.200	0.278	0.187
visibility_a3	0.509*	0.272	0.509*	0.274	0.375**	0.163	0.375*	0.202
visibility_a4	0.232	0.184	0.232*	0.141	0.031	0.109	0.031	0.113
country_a2	0.441**	0.175	0.441***	0.168	0.162	0.122	0.162	0.141
country_a3	0.652***	0.177	0.652***	0.169	0.314**	0.123	0.314**	0.142
country_a4	0.770***	0.214	0.770***	0.191	0.250*	0.137	0.250*	0.148
Observations	1173				1173			
Pseudo R^2	0.028							

***, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Table 24: Negative binomial model regression (NBM), negative binomial model regression with robust standard errors, zero-inflated negative binomial model regression (ZINB) and zero-inflated negative binomial model regression with robust standard errors of model 3. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_other_low_cost_flight_carriers*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Model 3	**NBM**		**NBM robust standard errors**		**ZINB**		**ZINB robust standard errors**	
Variable	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**
Ryanair_other_low_cost_flight_carriers	0.959*	0.494	0.959***	0.365	0.325	0.291	0.325	0.295
time_sched_d	0.052***	0.020	0.052***	0.017	0.028**	0.012	0.028**	0.014
type_airport_a	0.928**	0.387	0.928***	0.287	0.665***	0.235	0.665***	0.242
date	-1.890***	0.320	-1.890***	0.217	-1.056***	0.168	-1.056***	0.167
log_distance	-0.477	0.404	-0.477	0.337	-0.258	0.273	-0.258	0.268
visibility_a1	1.888	2.150	1.888***	0.246	1.069	0.941	1.069***	0.211
visibility_a2	0.882	0.631	0.882***	0.299	0.411	0.312	0.411	0.256
visibility_a3	0.608	0.415	0.608	0.379	0.382*	0.229	0.382	0.289
visibility_a4	0.121	0.279	0.121	0.178	-0.142	0.157	-0.142	0.161
country_a2	0.336	0.269	0.336	0.261	0.095	0.188	0.095	0.221
country_a3	0.730***	0.259	0.730***	0.249	0.371**	0.177	0.371*	0.219
country_a4	1.145***	0.329	1.145***	0.324	0.522**	0.212	0.522**	0.256
Observations	616				616			
Pseudo R^2	0.029							

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**

Table 25: Negative binomial model regression (NBM), negative binomial model regression with robust standard errors, zero-inflated negative binomial model regression (ZINB) and zero-inflated negative binomial model regression with robust standard errors of model 4. The dependent variable is *delay_a_minutes(2)* and the independent variable is *Ryanair_KLM*. *Visibilty_a* is a dummy variable with five categories. *Visibility_a5* is taken as base category. *Country_a* is also a dummy variable, with four categories. *Country_a1* is taken as base category.

Model 4	**NBM**		**NBM robust standard errors**		**ZINB**		**ZINB robust standard errors**	
Variable	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**	**Coeffi-cient**	**Std. err.**
Ryanair_KLM	0.809**	0.347	0.809***	0.298	0.522**	0.222	0.522**	0.260
time_sched_d	0.046***	0.014	0.046***	0.013	0.033***	0.009	0.033***	0.010
type_airport_a	0.940***	0.341	0.940***	0.276	0.713***	0.218	0.713***	0.234
date	-1.760***	0.192	-1.760***	0.151	-1.115***	0.116	-1.115***	0.125
log_distance	0.029	0.354	0.029	0.299	0.133	0.237	0.133	0.257
visibility_a1	1.174	1.030	1.174***	0.260	0.619	0.489	0.619**	0.247
visibility_a2	0.514	0.420	0.514**	0.246	0.190	0.232	0.190	0.217
visibility_a3	0.507*	0.302	0.507*	0.292	0.392**	0.182	0.392*	0.228
visibility_a4	0.296	0.199	0.296*	0.153	0.073	0.118	0.073	0.123
country_a2	0.373*	0.194	0.373*	0.195	0.095	0.138	0.095	0.163
country_a3	0.647***	0.195	0.647***	0.189	0.295**	0.138	0.295*	0.158
country_a4	0.704***	0.239	0.704***	0.205	0.219	0.151	0.219	0.160
Observations	916				916			
Pseudo R^2	0.031							

*****, ** and * stands consecutively for the 0.01, 0.05 and 0.10 significance level**