ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis Master Strategy Economics

The effect of quality on vertical integration in the airline industry

Name student: Rogier (R.J.G.) Beets

Student ID number: 476631

Abstract:

The economic literature has focused on transaction costs when studying vertical integration. A quality perspective can enhance understanding vertical integration. Quality can affect vertical integration due to the risk of reputation loss when outsourcing to independent firms. Using data on vertical integration in the domestic U.S. airline industry from 2016 to 2019, an instrumental variable method is applied with weather conditions as instrumental variable. A positive relationship between quality and the propensity to vertically integrate is found, which is most likely a causal effect. There is no significant difference in this effect between vertical integration by a major airline performing a flight itself and vertical integration by outsourcing to a regional airline.

Supervisor: H.P.G. Pennings

Second assessor: A.S. Bhaskarabhatla

Date final version: XXX

Table of contents

1. Introduction

When firms start to produce a product or offer a service, they have to decide how to obtain their supplies and how to distribute to their customers. Firms can produce parts of the vertical chain in house or buy them on the market. Towards the consumer, firms can distribute their products themselves or let a distributor do this. The in-house production and distribution are called vertical integration.

Vertical integration is a subject that has been studied for a long time in the economic literature. Already in 1937, Coase introduced the idea of transaction costs and discussed the boundaries of the firm. As Carlton (1979) interprets Coase's work, activities are performed within a firm when "its internal allocative function is superior to that of a market" (p. 190). This means that firms apply some form of vertical integration when they are better at performing the activity than the market. Maddigan (1981) applies Coase's work and thinks that the essence of vertical integration lies in the difference between market forces and management control to coordinate the optimal allocation of inputs.

The analysis of vertical integration has focused mainly on transaction costs, based on the Transaction Costs Economics of Williamson (1979). For example, Monteverde and Teece (1982) have studied the effect of transaction costs on vertical integration in the automotive industry. However, a quality perspective has not often been taken when studying vertical integration. Only Fernández-Olmos, Rosell-Martínez and Espitia-Escuer (2009) have applied a quality perspective, when studying vertical integration in the wine industry and find that quality can explain vertical integration as well as transaction costs. This suggests that more research is needed on the relationship between quality and vertical integration.

The airline industry sets a good environment to study this relationship, as major airlines can perform a flight themselves, outsource the flight to an owned regional airline or outsource the flight to an independent regional airline. Regional airlines have more unexperienced employees compared to major airlines, which results in lower operating costs for regional airlines, but also a lower quality (Forbes & Lederman, 2007). These lower operating costs are an incentive for major airlines to outsource to regional airlines, but the lower quality can be an incentive to vertically integrate.

Forbes and Lederman (2009) have previously studied vertical integration in the airline industry. They used a Transaction Cost Economics approach and found that vertical integration by outsourcing to an owned regional airline is more likely than outsourcing to an independent regional airline when adaptations in flight schedules occur more often due to adverse weather conditions. As Forbes and Lederman show, the presence of an exogenous variable as weather conditions that influences the behaviour of airlines makes the airline industry interesting for studying vertical integration. In addition, the airline industry also shows to be a good setting to study quality, as quality can objectively be measured by on-time performance (Van Reeven & Pennings, 2016). Data on on-time performance are available on a large scale. In addition, major airlines have a reputation of high quality, which can create an incentive for vertical integration.

Therefore, the following research question will be studied:
Does quality affect vertical integration in the airline industry?

This research will try to explain vertical integration by quality. In addition, the difference between vertical integration by a major airline performing a flight itself and vertical integration by outsourcing to an owned regional airline will be studied. Importantly, quality is an endogenous variable in the airline industry, as vertical integration increases quality (Forbes & Lederman, 2010). Therefore, an instrumental variable (IV) approach will be taken in the empirical analysis with weather characteristics as exogenous variable.

The research question is chosen in order to further study the interplay between quality and vertical integration, especially in the airline industry. The effect of quality on vertical integration has not been studied often yet. However, there is evidence of a relationship between vertical integration and quality. For example, Arocena (2008) finds evidence of an increase in quality after vertical integration in the electricity industry, due to higher efficiency. Contrary, Short and Ho (2020) only find a small increase in certain aspects of quality after vertical integration in the health industry. On the other hand, the literature previously discussed has also found evidence that quality affects vertical integration, which is related to reputation and moral hazard (Fernández-Olmos et al., 2009). Regarding the airline industry, Forbes and Lederman (2010) have found that vertical integration positively affects performance and quality. Seen in the light that quality also can affect vertical integration as argued previously, the proposed research will further describe the interplay between quality and vertical integration in the airline industry. This will contribute to the research of Forbes and Lederman (2009) and will provide a good example of the interplay between quality and vertical integration.

In order to take into account that vertical integration can affect quality an instrumental variable approach is used, as follows. The first stage regresses on-time performance on weather characteristics as instrumental variable, while controlling for flight, airport, firm and time characteristics. Then, the second stage regresses vertical integration on the fitted values, while controlling for the same variables. Weather conditions, measured by precipitation, will serve as instrumental variable of quality, as weather conditions affect the difficulty of a route. If weather conditions are more adverse, delays are more likely (FAA, 2021). Hence, the quality of a flight will decrease.

The scientific relevance of this research is that vertical integration has mostly been studied from the perspective that contracts are imperfect. As Fernández-Olmos et al. (2009) mention, a quality perspective is applied less often. Furthermore, the proposed research will show the interplay between quality and vertical integration by building upon the research of Forbes and Lederman (2010). Additionally, weather conditions are clearly exogenous and therefore can provide valid scientific results.

This research will study vertical integration with a different approach than Forbes and Lederman (2009). They also study vertical integration in the airline industry. However, they look at the choice of a major airline whether to outsource to an owned regional airline or to an independent regional airline. As both types of airlines have unexperienced employees and therefore lower quality, the effect of quality on vertical integration cannot be studied. By integrating vertical integration by a major airline performing a flight itself in the choice how to perform a flight, the effect of quality can be studied, as major airlines have different characteristics compared to regional airlines, such as employees and reputation.

The societal relevance of this research is that the airline industry currently is in a difficult situation due to the Covid-19 pandemic. New insights on the role of quality in the airline industry can aid in resolving this situation, as quality is a choice variable for customers, which will be discussed in Section 2.4. Additionally, the Covid-19 pandemic has presented issues to supply lines and therefore highlights the importance of firm boundaries. Furthermore, regional airlines could learn whether quality affects the integration decision of major airlines. For policymakers this research can also be of interest. The United States (U.S) Congress has been interested in improving the level of quality in the airline industry for a long time (Mazzeo, 2003). The results of this study on the interplay between quality and vertical integration can offer new insights to policymakers on how to improve the level of quality in the airline industry. In addition, the U.S. airline industry is known to need financial support from the government, which has all sorts of negative consequences such as lower pensions for workers (Goetz & Vowles, 2009; Blair, 2003). Vertical integration affects operational performance by airlines and financial performance in general (Forbes & Lederman, 2010; D'Aveni & Ravenscraft, 1994). Therefore, policymakers should be interested in results on what affects vertical integration in the airline industry.

The main result of this study is that there is a positive relationship between quality and the propensity to vertically integrate. This relationship is most likely a causal effect. There is no significant difference in this effect between vertical integration by a major airline performing a flight itself and vertical integration by outsourcing to an owned regional airline.

This paper is organised as follows. First, the theoretical framework in Section 2 will describe the existing literature on vertical integration, quality and the airline industry. Based on this literature it will be argued that quality positively affects vertical integration in the airline industry. Two hypotheses will be formulated. After this, the data and methodology will be described in Section 3. A large dataset on all flights in the domestic U.S. airline industry and the delays of these flights will be used. To establish a causal effect, an instrumental variable approach will be applied with weather as exogenous variable. Section 4 will discuss the results of this empirical analysis. At last in Section 5, the results will be discussed and a conclusion will be drawn on the effect of quality on vertical integration in the airline industry.

2. Theoretical framework

This section describes how an effect of quality on vertical integration in the airline industry fits within the economic literature. First, vertical integration in general is described. Then, it is explained how vertical integration occurs in the airline industry. Next, quality and its relationship with economic activities and strategy are described. Thereafter, quality is connected to vertical integration in general. At last, it is discussed how quality can affect vertical integration in the airline industry.

2.1 Vertical integration

Vertical integration is the integration of business activities up and down the value chain in one firm. Typically, vertical integration is framed as the make-or-buy decision. This entails that firms have to decide whether to buy an input or make the input themselves. However, vertical integration also applies to services and integration towards the customer by for example integrating distribution channels.

Different theories try to explain why firms use vertical integration. The most discussed theory is Transaction Costs Economics, which was developed by Williamson (1979). He starts from the notion that firms invest in a specific relationship with a supplier or distributor, which represents a value. However, contracts are incomplete, in the sense that contracts cannot incorporate all possible future situations or conflicts. This creates a risk for the value of the relationship and the investment, as each new situation creates the opportunity for one party to capture more rents. This opportunistic behaviour is a risk for the other firm, because the firms have to renegotiate to incorporate the new situation into the contract, which leads to transaction costs. To reduce the risk of opportunistic behaviour and avoid the costs of a renegotiations, firms integrate activities. Williamson argues that vertical integration is most likely when a recurring transaction involves transaction-specific investments and is made under uncertainty.

The Transaction Costs Economics of Williamson (1979) explains vertical integration as an action to avoid higher transaction costs. Another important theory on vertical integration is the Property Rights Theory of Grossman and Hart (1986). Grossman and Hart assume contracts to be incomplete, as Williamson (1979). However, the Property Rights Theory explains vertical integration as an action to optimally distribute investment incentives, instead of minimising transaction costs. When a transaction takes places, some property rights are transferred to another party. However, non-specified property rights, called residual rights, are not transferred, as contracts are incomplete. These residual rights will determine the distribution of the surplus. When new unexpected situations arise, the residual rights will determine who has the right to capture (a part of) the surplus. As the distribution of the ex post surplus between both parties is determined by the ownership distribution, ownership will also determine ex ante investment incentives. When a party expects to receive less of the surplus as it does not own the residual rights, this party will underinvest. This is the case for the party without ownership over the residual rights, whereas the owner of the residual rights has an incentive to overinvest. This distortion takes place regardless whether the transaction takes place on the market or within the firm, regardless whether one party is an independent party or a division or employee within a firm.

The distortion of investment incentives and the following suboptimal investment outcome, leads to a loss in surplus. Firms will choose to integrate the transaction and activity when the loss of surplus is smaller than when the transaction takes place on the market. Grossman and Hart (1986) state that vertical integration will take place when the investment of one party is more important than the investment of the other party.

It is important to notice that integration of an activity also can lead to more costs. Bureaucracy costs can arise when an activity is carried out within a firm, as a larger organisation and more hierarchy levels can reduce the efficiency of control over the organisation and miscommunications (Mahoney, 1992). In addition, a lack of market pressure can lead to bureaucracy costs, as profit incentives are lower between firm divisions than independent firms (Mahoney, 1992). D'Aveni and Ravenscraft (1994) find that vertical integration mostly increases bureaucracy costs due to a lack of market pressure and not due to size or hierarchy. These studies show that vertical integration can also lead to extra costs and not only reduce the costs of opportunistic behaviour. This means that firms only decide to vertically integrate when the costs avoided by integration are higher than the bureaucracy costs that will emerge due to the integration.

2.2 Vertical integration and quality

Much of the economic literature on vertical integration is based on Transaction Costs Economics (Whinston, 2001). Empirical studies have found strong support for the propositions of the Transaction Costs Economics and mostly for the relationship between asset specificity and vertical integration. As will be discussed later on, Transaction Costs Economics has also been used to explain vertical integration in the airline industry.

However, vertical integration might also occur due to other reasons. One of these reasons could be quality. Harrigan (1984) states that integration or quasi-integration can be performed for quality control, as full integration gives the firm tight control over the quality of the product or service. Control guarantees a constant level of quality and that the quality of the product or service is consistent with the image and reputation of the firm.

According to the theoretical and mathematical analysis of Lin, Parlaktürk and Swaminathan (2014), backward integration is preferred when product quality is an important factor in the market, because vertical integration gives the firm full control over the decision at which quality level to produce. As the integrating firm can then decide at which quality level to produce, can invest in quality and enjoys the benefits of higher quality, vertical integration can be beneficial. This effect increases when consumers are more sensitive to quality (Lin et al., 2014). So, when quality is an important factor for customers, vertical integration is more likely in order to control quality.

A relationship between quality and vertical integration has also been supported empirically. Fernández-Olmos et al. (2009) find that in the wine industry both Transaction Costs Economics and product quality can explain vertical integration. The researchers investigate vertical integration from a quality perspective as previous research has focused mainly on a Transaction Costs Economics approach. Wineries can fully or partially integrate the production of grapes, as it is possible to partially buy grapes from other firms. Fernández-Olmos et al. (2009) find that producers of high-quality wines are more likely to integrate grape production, because wineries with a reputation of high-quality wines have a competitive advantage in this reputation. However, the quality of grapes is hard to measure by the winery, which means that wineries cannot control independent grape producers. Therefore, wineries will risk moral hazard when they outsource grape production. To avoid losing reputation, grapes must be produced by the winery itself. Therefore, vertical integration is more likely for wineries with high-quality wines. This hypothesis is empirically supported, by defining high-quality wines as wines that have aged for more than three years and wines selected by experts as high-profile wines. These wines are known for their quality and consequently risk moral hazard when outsourcing (Fernández-Olmos et al., 2009).

The empirical findings of Fernández-Olmos et al. (2009) relate to the mathematical findings of Lin et al. (2014). According to Lin et al. (2014), vertical integration is more likely when quality is an important factor for customers, which is the case in the wine industry, where grape quality is important for the quality of the wine. Wineries compete with each other on quality, so grape quality is essential to wineries (Fernández-Olmos et al., 2009). Therefore, the mathematical findings of Lin et al. (2014) are supported by the empirical findings of Fernández-Olmos et al. (2009).

The argumentation based on a competitive advantage due to high quality and a reputation of high quality has also been used to explain vertical integration in the health care sector by Coles and Hesterly (1998). In the health care industry quality is important for the reputation of hospitals, because quality is an important factor for competition in this industry. Hospitals with a reputation of high quality have a competitive advantage over other firms. In addition, these hospitals can more easily attract higher skilled employees, which can lead to attracting more patients and less litigation costs. Litigation costs can follow from failed medical procedures, which are less likely for higher skilled employees. Similar to the wine industry, quality is hard to measure in the health care industry. When a procedure fails, it is hard to establish whether this is the fault of the health care worker or a solely medical occurrence. Therefore, when an independent health care worker would make a mistake, the hospital probably would pay (part of) the litigation costs following from this mistake, even though these costs could be shifted to the independent worker. In addition, the reputation of the hospital would suffer from this mistake. Therefore, hospitals with a reputation of high quality would be more likely to integrate the performance of a medical procedure in order to preserve their reputation and avoid litigation costs.

This reasoning is empirically supported. Services that are more likely to affect the quality of the provided health care and consequently more likely to affect reputation, are more likely to be integrated than services with less risk of affecting the health care. Even when other reasons for vertical integration resulting from transaction costs are inclined towards outsourcing, vertical integration is still more likely due to the risk of reputation loss (Coles & Hesterly, 1998). This shows that quality and reputation are very strong determinants of vertical integration in the health care industry.

These studies show that quality can affect vertical integration. Besides transaction costs, quality can also create an incentive for firms to produce a product or perform a service themselves.

2.3 Vertical integration in the U.S. airline industry

In order to describe how vertical integration occurs in the U.S. airline industry, first three different types of airlines will be described. The U.S. airline industry is characterised by the presence of major airlines, low-cost airlines and regional airlines. Major airlines make up 50.0% of the domestic market, whereas low-cost airlines make up 39.1% of the market (T4, 2021). Regional airlines account for the last 10.9%, of which 2.9 percentage points are accounted for by the largest regional airline SkyWest Airlines. To understand how these airlines behave and compete, we have to look at the history of the airline industry.

Major airlines

Major airlines are the original airlines in the U.S. airline industry. Therefore, they are also called legacy airlines. Nowadays, major airlines offer flights with high service quality against high prices and often operate adjacent activities themselves, such as the food service and luggage handling.

Around 1930 the airline industry was unstable (Goetz & Vowles, 2009). To solve this, the U.S. government decided to regulate the industry. The Civil Aeronautics Board regulated the entry and exit of airlines on specific routes as well as prices. A number of airlines existed and were awarded a mix of small and large routes to compensate unprofitable routes. As prices were regulated, airlines focused on service quality. Therefore, these airlines are also called full-service airlines, as they still offer high quality flights.

The regulation led to market inefficiencies, so the government deregulated the industry from 1978 on. Entry, exit and prices were no longer regulated by the Civil Aeronautics Board (CAB). In general, this deregulation led to lower prices and more efficiency. However, less-travelled routes faced higher prices and lower levels of service. In addition, financial performance of airlines decreased and bankruptcies occurred more often (Goetz & Vowles, 2009). Major airlines started to use a hub-and-spoke network instead of direct flights (Borenstein, 1992). A hub-and-spoke network entails that airlines operate from a central airport. Passengers fly to this central airport and then transfer onto a different flight towards their final destination.

Deregulation also led to the entry of low-cost airlines, as will be described below (Cento, 2009). As a reaction to the entry of low-cost airlines and the following decrease in prices, the major airlines had to improve their service and quality. In addition to service improvements, costs also had to be decreased in order to close the costs efficiency gap between the major airlines and low-cost airlines.

Low-cost airlines

Besides the described changes, low-cost airlines emerged due to the deregulation of entry and exit of firms. As prices were no longer regulated, airlines could compete on price. Low-cost airlines offer flights against low prices and lower quality (Borenstein, 1992). Low-cost airlines do not operate a hub-and-spoke network but a point-to-point network and usually only offer short-haul flights (Hunter, 2006). In addition, low-cost airlines outsource ground services, such as luggage handling, and use secondary airports to fly to and from, for example London Gatwick instead of London Heathrow. Low-cost airlines have large competitive effects when entering a new route (Brueckner, Lee & Singer, 2013).

Regional airlines

Besides major airlines and low-cost airlines, the U.S. domestic airline industry is also characterised by the existence of regional airlines. Regional airlines did not emerge after the deregulation, as low-cost airlines, but during the period of regulation. Already in 1944, the CAB allowed so-called local service airlines to enter the market, as response to increasing demand (Forbes & Lederman, 2007). These feeder airlines flew on routes to small communities, without competing with the already existing airlines. In addition, in 1949 commuter airlines were allowed to enter the airline market. These airlines flew on irregular routes, like taxis. However, commuter airlines were only allowed to fly small aircrafts.

Because the routes to small communities were not profitable, mainly the local service airlines had to be subsidised. When subsidies decreased from 1960 on, commuter airlines took over the role of the local service airlines to serve local communities. From 1964 commuter airlines started codeshare relationships with the major airlines. Commuter airlines could then serve the local communities for the major airlines. After the deregulation in 1978, commuter airlines could replace major airlines and local service airlines at more routes and fly with larger airplanes. As major airlines chose to use a hub-and-spoke network, codeshare relationships became more prevalent, as passengers from local communities preferred easy connections. Commuter airlines, which would turn into regional airlines, would feed passengers into the network of the major airlines (Forbes & Lederman, 2007).

Nowadays, regional airlines operate under codeshare agreements with one or multiple major airlines (Forbes & Lederman, 2007). In general, regional airlines do not fly under their own code. Major airlines subcontract the performance of flights to regional airlines, which means that regional airlines physically perform the flight and use the brand of the major airline. Usually, short and low-density routes are subcontracted to regional airlines, which is the reason that regional airlines usually operate smaller aircrafts than major airlines.

Codeshare agreements can be made with independent regional airlines or owned regional airlines. Independent regional airlines usually own their own aircrafts and have their own employees. On the other hand, the aircrafts and employees of owned regional airlines are usually part of the major's aircraft fleet and workforce. This might lead to a reduced cost benefit of owned regional airlines, compared to independent regional airlines. Employees of owned regional airlines might seek wages that are similar to the wages of employees in the same position at the major airline (Forbes & Lederman, 2007).

Vertical integration between major airlines and regional airlines

In the U.S. domestic airline industry major airlines let regional airlines perform some routes (Forbes & Lederman, 2009). Essentially, these flights are "produced" outside the firm by outsourcing the flight to regional airlines. The reason for this is the cost advantage of regional airlines, which relates to one of main characteristics of the U.S. airline industry, namely labour mobility. Major airlines pay their employees based on seniority. Additionally, major airlines face labour unions with large bargaining power. On the other hand, regional airlines usually employ less experienced employees with a lower wage, because many young pilots are willing to work for a lower salary in order to gain experience and eventually start working for a major airline. In addition, not all regions have labour unions. Therefore, regional airlines are often not bound by the high salary demands of labour unions and consequently can operate at lower costs. Major airlines benefit from this cost advantage by outsourcing the performance of the flight to regionals (Forbes & Lederman, 2009).

Outsourcing of the performance of the flight can take place in two manners, namely by outsourcing to an owned regional airline or to an independent regional airline. Forbes and Lederman (2009) have found that major airlines more often choose to let an owned regional airline perform the flight when adaptations in the flight schedule occur more often due to adverse weather or when the costs of adaptations are higher due to a higher level of integration into the network of the major airline. The likelihood of adaptations relates to complexity as subgroup of asset-specificity of the Transaction Costs Economics (Lafontaine & Slade, 2007).

The study of Forbes and Lederman (2009) separates the decision to outsource from the decision whether to outsource either to an owned regional airline or to an independent regional airline. Using weather conditions as measurement of the likelihood of adaptations, they obtain valid results with an exogenous variable.

2.4 Quality in the airline industry

Quality is an important factor in the airline industry. When competition increased after the deregulation of the airline industry, quality became more important for airlines due to competitive pressure (Ostrowski, O'Brien & Gordon, 1993). In the US, Congress and other policy makers have been involved with the service level and Congress has adopted legislation to increase the service level (Mazzeo, 2003). Mazzeo studies whether competition affects quality, measured by on-time performance. The results of his research shows that quality is an important and strategic variable for airlines, as follows.

If delays are more likely, passengers will choose to fly with another airline. However, if there is no alternative flight, the implications of delays will be less severe, as customer have no other flight to change to. The absence of an alternative can be due to competition, but also frequent flyer programs or for other reasons. This reasoning has been supported by Suzuki (2000). Suzuki estimated a model which shows that passengers who experience a delay are more likely to switch to another airline for their next flight than passengers who did not experience a delay. This result also reveals that quality affects firm performance through passengers' experience.

In the same line of reasoning, Mazzeo (2003) found different results that show that quality is a strategic variable for airlines. If the destination of a route is a hub of the airline, both the probability of a delay and the magnitude of the delay decreases, probably because of connecting passengers. If passengers miss their connecting flight, the delay for these passengers increases to a great extent, which leads to more frustration of the passenger and higher compensation costs.

Regarding the relationship between competition and quality, different results are relevant. If only one airline flies directly on a route, delays are more likely. This is the same if the Herfindahl-

Hirschman-index, as measurement for market concentration, increases. This shows that delays and therefore quality are impacted by competition, probably because less competition creates less incentives for airlines to reduce delays.

Most interesting, Mazzeo (2003) finds evidence that delays are more likely to occur on less competitive routes that originate from a competitive hub than on less competitive routes originating from a less competitive hub. This suggests that airlines use staff or equipment from less competitive route to prevent delays on competitive routes from the same hub.

These different results show that airlines regard quality as a strategic variable. Quality is not set in advance at a constant level relating to the fare. In contrast, the quality level is used to maximise total profit by setting the quality each flight again in light of the other flights of the airline. Airlines deliberately choose which routes to serve with high quality and which routes with low quality. Airlines even set quality lower for one flight to keep quality up for another flight, which is in risk of a low level of quality.

Quality is also an important factor for customers, which is found in results on the relationship between quality and consumer loyalty. Ostrowski et al. (1993) find evidence that quality is positively related to consumer loyalty in the airline industry. Passengers' long-term image of the airline is most important for creating consumer loyalty. This is important for airlines, as consumer loyalty is important for airlines to retain market share (Ostrowski et al., 1993). Zeithaml, Berry and Parasuraman (1996) find that consumer loyalty improves profitability, as loyal customers require less marketing than attracting new customers. In addition, loyal customers are more likely to spend more on additional services and tell others about their good experiences with the firm. These studies show that quality is an important factor for customers in the airline industry.

The discussed literature shows that quality is a strategic variable for airlines, a choice variable for customers and of importance to policy makers.

2.5 Vertical integration and quality in the airline industry

Vertical integration takes place in the U.S. airline industry in the form of major airlines performing flights or outsourcing the flights to owned regional airlines instead of outsourcing to independent regional airlines. Vertical integration can be explained by, among other reasons, transaction costs and property rights. As Forbes and Lederman (2009) showed, the Transaction Costs Economics applies to vertical integration in the U.S. airline industry. The increased likelihood of adaptations in flight schedules when outsourcing and the resulting transaction costs move major airlines to outsource to owned regional airlines instead of independent regional airlines.

As described, quality can also be related to vertical integration. Theoretical and empirical studies have found evidence of a relationship between quality and vertical integration. Most interesting is the study of Fernández-Olmos et al. (2009). They find that quality is positively related to vertical integration, due to the risk of reputation loss when outsourcing. Lin et al. (2014) mathematically find that vertical integration is more likely in industries where quality is an important factor.

As described, quality is an important factor in the U.S. airline industry. Policymakers focus on quality and want to improve quality. Customers find quality an important factor for their loyalty to airlines, which is based on the reputation of an airline. For airlines, quality is also important, as retaining loyal customers by high quality can improve profitability (Zeithaml et al., 1996). Airlines regard quality as a strategic variable, which is shown by the fact that airlines provide high quality on competitive routes and lower quality on routes with no competition (Mazzeo, 2003). This is also shown by the fact that airlines deliberately set a lower quality level at certain moments on certain routes in order to provide high quality on other routes.

So, this literature shows that quality can be related to vertical integration, specifically in industries where quality is important, which is the case in the U.S. airline industry. Therefore, quality might also affect vertical integration in the airline industry.

As described, the low level of labour mobility in the U.S. airline industry means that experienced pilots are employed by major airlines for a high wage and inexperienced pilots are employed by regional airlines for a low wage. The labour cost advantage of regional airlines over major airlines creates an incentive for major airlines to outsource their flights to regional airlines. However, quality can create an incentive for vertical integration due to the risk of reputation loss.

As more experienced pilots are employed by major airlines, the quality of flights will be higher for these airlines. Therefore, major airlines have a reputation of high quality, which creates a competitive advantage. The competitive advantage entails that customers will be more loyal to the airlines, which improves profitability. On the other hand, for regional airlines, the quality of flights will be lower. This means that the probability of lower quality is higher when outsourcing compared to when major airlines perform flights themselves. This higher probability creates the risk of reputation loss, when outsourcing. To avoid reputation loss major airlines can perform high quality routes themselves, as these routes have a high risk of a loss of quality when outsourcing. This reasoning relates to the research of Fernández-Olmos et al. (2009).

In addition to reputation loss, low quality could also lead to extra costs. This relates to the previously discussed research of Coles and Hesterly (1998), which showed that outsourcing is less attractive when the independent party can avoid bearing the costs of low quality and shift these costs to the outsourcing party. This can be the same in the airline industry. The lower quality provided by regional airlines could lead to delays. These delays might have to be compensated. Regional airlines can argue that these delays are not their fault and transfer these costs over to the major airline. In addition, regional airlines can foresee these extra costs and include these costs in the contract with the major airline. This increases the costs of outsourcing and thus decreases the attractiveness of outsourcing. Major airlines are then more attracted to performing routes on which delays do not occur themselves, to avoid the extra costs.

Therefore, major airlines would be more inclined to perform high quality routes themselves to protect reputation and avoid extra costs, instead of outsourcing to regional airlines. However, this can be different for owned regional airlines. As major airlines own these airlines, they have the possibilities and power to adjust the performance of owned regional airlines. Therefore, the risk of reputation loss and of extra costs would be lower. If an owned regional airline would create a risk of reputation loss, the major airline could step in and change the way the regional airline operates, consequently reducing the risk of reputation loss. Therefore, the risk of reputation loss and extra costs applies only to outsourcing to an independent regional airline.

Therefore, the previous discussed empirical findings on the effect of quality on vertical integration can also apply to vertical integration in the airline industry. The literature on quality and vertical integration has found a positive relationship between those aspects. This results in the following hypothesis. Major airlines are more likely to perform high quality flights themselves or vertically integrate by outsourcing to owned regional airlines, instead of outsourcing to independent regional airlines. This reasoning is based on the risk of reputation loss and of extra costs in the case of outsourcing to independent regional airlines. The following hypothesis follows from this argument:

Hypothesis 1: There is a positive causal effect of quality on the propensity to vertically integrate.

This hypothesis is based on the reasoning that high quality flights have more risk of reputation loss than low quality flights. This increases the attractiveness of vertical integration. Therefore, a positive causal effect of quality on the propensity to integrate is hypothesised.

The first hypothesis regards vertical integration of a major airline by performing a flight itself as similar to vertical integration by outsourcing to an owned regional airline. However, there might be a difference between the two forms of vertical integration by major airlines. The characteristics of a major airline remain different of the characteristics of an owned regional airline, even though ownership grants power to change performance. Owned regional airlines might be more alike independent regional airlines than major airlines, as owned regional airlines have the same unexperienced employees as independent regional airlines. The resulting lower costs are the essential benefit of owned regional airlines over major airlines why major airlines decide to use owned regional airlines. However, due to these unexperienced employees owned regional airlines probably remain to offer lower quality than major airlines. Even though major airlines can control owned regional airlines, the difference in experience of pilots might not be fully mitigated. Therefore, some of the risk of reputation loss and of extra costs remains when applying vertical integration by outsourcing to owned regional airlines.

This means that the hypothesised effect under hypothesis 1 can be different for the two different types of vertical integration. As the risk of reputation loss and of extra costs might partially remain, vertical integration by outsourcing to owned regional airlines might be less attractive under circumstances than vertical integration by major airlines by performing flights themselves. These circumstances could be quality. For routes with very high quality the partially remaining risk of reputation loss might still induce major airlines to perform flights themselves.

Therefore, the effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to owned regional airlines. Independent regional airlines are more alike owned regional airlines, which means that the effect of quality might be weaker. This leads to the second hypothesis:

Hypothesis 2: The positive causal effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to owned regional airlines compared to vertical integration of major airlines by performing flights themselves.

3. Data & Methodology

This section describes the data and methodology that are used for the empirical analysis of the two hypotheses. The empirical analysis is based on an instrumental variable approach. First, the data sources for all variables are set out. Then, descriptive statistics are reported for the data. At last, the methodology for the empirical analysis is described and explained.

3.1 Data sources

The dataset is primarily based on data of the U.S. Bureau of Transport Statistics on delays of all domestic flights in the U.S. from 2016 to 2019 (Bureau of Transportation Statistics, n.d.). Airlines that account for at least 1% of domestic passenger revenue in the U.S. are required to report their flight data to this Bureau. The observations in the initial dataset are at flight level. Codeshare flights are only taken into account once and only direct flights will be examined. The sample consists in total of eighteen airlines. These include five major airlines, of which four use outsourcing to regional airlines, seven regional airlines, of which three are owned by a major airline and four are independent, and six low-cost airlines, which do not outsource the performance of flights. As a robustness check, a sample with only the major airlines that use outsourcing to regional airlines and the regional airlines will be used. This can show whether the obtained results are clear and not obscured with major airlines or low-cost airlines that never outsource. The data are available for a long period. So, to exploit variation over time, data will be collected from 2016 to 2019. 2020 will not be used, as the Covid-19 pandemic started to have an impact on the airline industry in 2020, which might obscure the results.

The initial dataset at flight level consists of 25,927,762 observations. To make the data better workable, the data are converted to month level. As will be explained, the independent variable is the arrival delay of a flight. This variable is averaged over the month to obtain the average arrival delay for a flight per month. After converting the initial dataset, the final dataset consists of 385,567 observations.

3.1.1 Dependent variable

The dependent variable in the empirical analysis of hypothesis 1 is a dummy indicating whether the performing airline is a major airline, a low-cost airline or an owned regional airline on one hand or an independent regional airline on the other hand. The dummy takes one as value if a major airline or owned regional airline performs the flight, thus the major airline has integrated the flight. The dummy takes zero as value if an independent regional airline performs the flight, which shows that a major airline has outsourced the flight. All flights performed by a regional airline in the data are outsourced flights, which is not a problem, as regional airlines do not fly under their own code (Forbes & Lederman, 2007).

For the second hypothesis, a different dependent variable will be used, which will be a categorical variable. The variable indicates whether the flight is integrated by a major airline or low-cost airline performing the flight itself, integrated by outsourcing to an owned regional airline or outsourced to an independent regional airline.

Data on which airlines are major airlines, owned regional airlines and independent regional airlines are derived from the annual reports from the Regional Airline Association from the relevant years (RAA, n.d.). The Regional Airline Association reports annually which regional airlines perform flights for which major airlines and whether these regional airlines are owned by a major airline or operate independently.

3.1.2 Main explanatory variable

The main explanatory variable is quality. Quality is often measured in the airline industry by on-time performance (Van Reeven & Pennings, 2016). The empirical analysis will take the same approach. The data on on-time performance are derived from the dataset of the Bureau of Transport Statistics, as previously described. Based on the hypotheses, it is expected that on-time performance has a positive effect on vertical integration. If routes are more often performed on time, quality is higher, as the risk of reputation loss is larger for the major airlines. Therefore, major airlines would be more likely to perform a flight themselves instead of outsourcing the flight.

On-time performance is measured by the arrival delay of a flight in minutes. Delay is an inverse measure of on-time performance. The arrival delay is used instead of the departure delay to capture the on-time performance of the whole flight. If the arrival delay was missing, it was assumed that there was no delay. This was assumed because the minimum observed value is equal to one, which means that every observed flight would have at least a minute of delay. It appears to be reasonable that there are some flights with no delay.

The distribution of the variable is presented in Figure A.1 in the Appendix. There are more observations with an observed delay of zero minutes than expected following a normal distribution. In addition, there are outliers, as the maximum value is 1590. To account for this skewed distribution, the logarithmic value of the delay plus one is used as main explanatory variable. One minute is added to each delay in order to account for the fact that there are observations with zero minutes of delay, which cannot be transformed into a logarithmic value.

Quality is an endogenous variable (Forbes & Lederman, 2010). Therefore, an instrumental variable (IV) approach will be taken with weather characteristics as exogenous variable.

3.1.3 Instrumental variable

The instrumental variable will be weather conditions. Weather conditions are measured by average daily precipitation in inches per month. Mazzeo (2003) found that precipitation and snowfall affect on-time performance. As discussed, if weather conditions are more adverse, for example due to high levels of precipitation, delays are more likely and on-time performance will be worse (FAA, 2021). Therefore, it is expected that precipitation has a negative effect on on-time performance.

The data on weather conditions will be collected from a database on historic weather data from the National Oceanic and Atmospheric Administration (NOAA) (NOAA, n.d.). The NOAA provides data on all observations of weather stations across the U.S. For each airport on which an airline has flown in the dataset, a corresponding weather station is searched manually. If the airport has a weather station itself, this weather station is chosen. This is the case for most observations, which makes the weather data highly reliable. If there is not a weather station on the airport, a weather station in the same town or county as the airport is chosen. Weather data on in total 328 stations were collected.

The weather observations are reported per hour per day. The average of the weather conditions per month is determined per station to work with the data, as previously done with the delay data. In total the weather data consist of 17,087,564 observations. After transforming the data to month level, the data consist of 15,375 observations. If the average precipitation per month was missing, it was assumed that there was no precipitation in this month. This was assumed because the observed minimum value is equal to one, which means that every observed station had at least one inch of precipitation in a month. It appears to be realistic that there are some stations with no precipitation in a given month.

The distribution of this variable is presented in Figure A.2 in the Appendix, which shows a skewed distribution and not a normal distribution. To account for this, the logarithmic value of the precipitation plus one is used as instrumental variable. A value of one is added to each observation in order to account for the fact that there are observations with zero minutes of delay, which cannot be transformed into a logarithmic value.

3.1.4 Control variables

The choice for control variables mainly follows Forbes and Lederman (2009). They find that the distance between airports, the population in the cities of both airports and a hub being present at one of the endpoints of a flight significantly affect the decision whether to use a regional airline or not. These variables reflect the importance of a flight for the network of the airline. Therefore, these variables are used as control variables.

The populations in the cities of both endpoint airports indicate the demand for a flight and therefore affect vertical integration, because if there is more demand, larger aircrafts are needed. Regional airlines usually operate with smaller aircrafts, so these airlines are less suited for routes with large populations at the endpoints. The population data are provided by the United States Census Bureau (United States Census Bureau, n.d.). Per flight, the population of the cities at both endpoints is used. Some airports are located near to multiple cities or a metropolitan area. For these airports, the sum of the population of these cities is used as value for the population. The dataset consists of 317 cities. For each route two variables on population are used, one for the population of the city of origin and one for the population of the city of destination. The histogram in Figure A.3 in the Appendix shows that the distribution of the population variables does not follow a normal distribution, but is skewed and has outliers. Therefore, the logarithmic value of the population variables is used.

The distance between both endpoints of a flight is also related to the type of aircrafts used by regional airlines. Smaller aircrafts usually have a smaller flight range, meaning that regional airlines cannot serve routes over a large distance. In addition, Forbes and Lederman (2009) mention that smaller aircrafts are more efficient on shorter routes. This would also mean that regional airlines are less likely to serve routes over a larger distance. The histogram in Figure A.4 in the Appendix shows that the distribution of the distance variables does not follow a normal distribution, but is skewed and has outliers. Therefore, the logarithmic value is used of the distance variable. The data on the distance between both endpoints are provided in the dataset of the Bureau of Transport Statistics on on-time performance and is measured in miles. There are 120 missing values. Therefore, the dataset consists of 385,447 observations.

Two dummies indicating whether the origin airport or the destination airport is a hub of the performing airline, are also included. Including this variable as control variable is not primarily based on reasoning, but on empirical findings. A hub can indicate the importance of a route and consequently create a reason for a major airline to serve the route itself. In addition, flights with a hub as an endpoint carry more transferring passengers (Forbes & Lederman, 2009). This means that the consequences of delays are larger when a hub is an endpoint. These larger consequences can create an incentive to minimise delays and therefore improve quality, which can be achieved by integrating.

On the other hand, regional airlines function as feeders for major airlines, as described in Section 2.3. Therefore, a hub as endpoint could also be a reason for major airlines to outsource a flight to a regional airline. Forbes and Lederman (2009) found that if one of the endpoints is a hub, some major airlines are more likely to outsource the flight to a regional airline. However, this is only the case for four out of the seven major airlines studied by Forbes and Lederman. Although it is not clear in what way a hub as endpoint affects the decision of a major airline whether to integrate or outsource a flight and the results of Forbes and Lederman (2009) also do not give a clear indication, it is important to include a hub variable as control variable. There is namely some evidence of an effect of a hub as endpoint on the choice whether to integrate or outsource.

The data on which airports are hubs for which major airlines are retrieved from the websites of the major airlines (Alaska Airlines, 2020; American Airlines, n.d.; Delta, 2021; Hawaiian Airlines, n.d.; United, n.d.). Two dummies are used to account for the effect of a hub on an endpoint, because the effect of a hub as the origin airport might be different from the effect of a hub as destination airport, as there will be more transferring passengers on a flight to a hub than from a hub. The dummies take one as value when one of the endpoints is a hub and zero if otherwise. For regional airlines, the hubs of the major airline that outsources the flight are regarded as the hubs of the regional airlines. Regional airlines do not have a hub of their own, but they function as feeders for the hubs of their major airlines. Therefore, there might be an effect of the hubs of the major airlines for the regional airlines. For example, a major airline could require more accurate on-time performance for flights to its hubs by regional airlines, as there are transferring passengers. This means that we can regard the hub of a major airline as the hub of a performing regional airline.

In addition to these variables, airport dummies are included to control for all other airport characteristics that can affect flight quality. If an airport is more difficult to depart from or land on, a high level of flight quality is more difficult to achieve. Besides airport characteristics, airline characteristics are also controlled for by airline dummies. At last, time characteristics are controlled for by time dummies.

3.2 Descriptive statistics

This section presents descriptive statistics on the data used. In addition, multicollinearity is investigated as possible concern.

Table 3.1 Descriptive statistics on the continuous and binary variables

Variables	Obs.	Mean	St. Dev.	Min	Max
Vertical integration (1 if true)	385,447	0.66	0.48	0.00	1.00
Arrival delay (log)	385,447	2.34	0.89	0.00	7.37
Precipitation at origin airport (log)	385,447	1.49	1.96	0.00	6.10
Precipitation at destination airport (log)	385,447	1.49	1.96	0.00	6.10
Population at origin airport (log)	385,447	13.02	1.53	5.79	15.95
Population at destination airport (log)	385,447	13.02	1.53	5.79	15.95
Hub at origin airport (1 if true)	385,447	0.25	0.43	0.00	1.00
Hub at destination airport (1 if true)	385,447	0.25	0.43	0.00	1.00

Distance (log)	385,447	6.56	0.74	3.43	8.54

Table 3.2 Descriptive statistics on the categorical variable on vertical integration for hypothesis 2

Category	Frequency	Percent	Cumulative
Outsourcing to an independent regional airline	252,447	27.28	27.28
Integration by outsourcing to an owned regional airline	27,835	7.22	34.50
Integration by performing	105,135	65.50	100.00
Total	385,447	100.00	

Table 3.1 presents descriptive statistics on the continuous and binary variables and Table 3.2 presents descriptive statistics on the categorical variable. The following statistics are predominantly relevant. 65.5% of the observed flights are performed by a major airline or low-cost airline and 34.5% by regional airlines. 7.2% of all the observed flights are performed by owned regional airlines and 27.3% by independent regional airlines. The average log of the arrival delay is 2.34. This translates to ten minutes. The maximum value of the log of the delay is 7.37, which is 26.5 hours. An average delay of ten minutes seems reasonable and does not rise questions on data issues. The maximum value of 26.5 hours highlights the importance of using the logarithmic value to account for outliers. The average daily precipitation per month is 1.49, which translates to 4 inches. 25% of the flights has been carried out with a hub as one of the endpoints. The log of the average distance between two endpoints is 6.56, which means 706 miles.

Table 3.3 Correlation matrix of all variables

		Variable								
Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)	Vertical integration (1 if true)	1.00								
(2)	Arrival delay (log)	-0.02 ***	1.00							
(3)	Precipitation at origin airport (log)	-0.01 **	-0.03 ***	1.00						
(4)	Precipitation at destination airport (log)	-0.01 ***	-0.02 ***	0.07 ***	1.00					
(5)	Population at origin airport (log)	0.06 ***	0.03 ***	0.14 ***	-0.02 ***	1.00				
(6)	Population at destination airport (log)	0.06 ***	0.06 ***	-0.02 ***	0.14 ***	-0.04 ***	1.00			

(7)	Hub at origin airport (1 if true)	-0.18***	-0.00	0.02***	-0.01***	0.30***	0.20***	1.00		
(8)	Hub at destination airport (1 if true)	-0.18***	0.01***	-0.02***	0.03***	-0.20***	0.30***	-0.17***	1.00	
(9)	Distance (log)	0.42***	-0.01***	0.04***	0.04***	0.14***	0.14***	-0.03***	-0.03***	1.00

Notes. *p < 0.05, ** p < 0.01, *** p < 0.001.

Table 3.2 presents a correlation matrix with all the variables. All the correlations are significant at an 1% significance level, except for the correlation between the variable on a hub at the origin airport and the arrival delay. Most of the correlation are below 10%, which suggests that there is only a small correlation between these variables. The following correlations stand out. The significant correlation between the arrival delay and vertical integration is negative and small. This suggests that higher delays and thus lower quality are associated to less vertical integration. This is in line with the expectations based on hypothesis 1. Average daily precipitation and arrival delays are significantly and negatively correlated, which suggests that higher average daily precipitation is associated to less or lower delays. This is contrary to the expectation that precipitation positively affects delays. However, this does not pose an immediate problem for the analysis, as a correlation does not account for control variables, such as the population at the endpoints. Still, this unexpected correlation shows that it is important to pay attention to the coefficient of precipitation on delays in the following regressions.

The correlations between the control variables and vertical integration do not rise concerns. The size of the population at an endpoint is slightly positively correlated to vertical integration, as expected. In line with the expectations is the positive and significant correlation between the distance variable and vertical integration, what means that regional airlines are less likely to serve routes over a larger distance. Furthermore interesting are the correlations between the hub variables and vertical integration. These correlations are significant and -18%. This means that a hub being present at one of the endpoints of a flight is negatively associated to vertical integration. This is in line with the results from Forbes and Lederman (2009) as mentioned in Section 3.1.4.

The correlation matrix shows mostly relatively small correlations. This means that there are no concerns towards multicollinearity. To further investigate this, two Variance Inflation Factor (VIF) tests are performed. The first test regresses the delay on the other variables. The second test regresses vertical integration on the delay and the other variables. The VIF scores are all very low and below 10 in both tests, which is a general rule of thumb (Curto & Pinto, 2011). The VIF tests and scores are presented in Table A.1 and Table A.2 in the Appendix. Therefore, we can conclude that there is no risk of multicollinearity.

3.3 Methodology

To study the effect of quality on vertical integration an instrumental variable approach is taken. Quality is an endogenous variable. Therefore, an IV approach with weather conditions as instrumental variable is used. This means that two models are used.

The model for the first stage is as follows:

1. $Delay_{fatod} = \beta_0 + \beta_1 * Average\ Precipitation\ at\ Origin\ airport_o + \beta_2 * Average\ Precipitation\ at\ Destination\ Airport_d + \beta_3 * Distance_{od} + \beta_4 * Population_o + \beta_5 * Population_d + \beta_6 * Hub_o + \beta_7 * Hub_d + Origin\ airport_o + Destination\ airport_d + Airline_a + Time_t + \varepsilon_{ict}$

The second stage model is as follows:

2. $\Pr\,(VI_{fatod} = 1) = \Phi(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t) + \varepsilon_{ict}$

Here, *VI* means vertical integration; $\widehat{D}elay$ means the predicted values for the delay; *f* indicates the specific flight; *a* indicates the operating airline; *t* indicates the date of the flight; *o* indicates the origin airport and *d* indicates the destination airport. *Origin airport* and *Destination airport* is a set of dummies to control for airport characteristics. The variable *Airline* is a set of dummies to control for characteristics of the operating airlines. The variable *Time* is a set of dummies to control for time effects.

The first stage consists of a pooled ordinary least squares (OLS) regression with delay, an inverse measure of on-time performance, as dependent variable and weather conditions as explanatory variables. The second stage consists of a probit regression with the vertical integration dummy as dependent variable and the predicted values of the delay as explanatory variable.

Hypothesis 1 is tested based on the coefficient for the predicted values of the delay, β_1 in regression 2. A positive effect of quality on the propensity to vertically integrate is hypothesised. As quality is measured by arrival delays and this is an inverse measure, higher delays will mean lower quality. This means that a negative effect of delays on the propensity to vertically integrate is hypothesised. Therefore, hypothesis 1 can be confirmed if the coefficient is significantly smaller than zero.

To test the second hypothesis, a different second stage is used. The first stage remains model 1. A multinomial logit model (MLM) is estimated with a categorical variable as dependent variable to test the difference in the effect of quality on vertical integration between owned regional airlines and independent regional airlines. The dependent variable indicates whether the flight has been performed by an independent regional airline, by an owned regional airline or by a major airline. In this multinomial logit model outsourcing to an independent regional airline is used as the baseline category. The model for the second stage is as follows:

3. $\Pr(VI = Self-performed_{fatod} = \exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{self-performed} / (\exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{self-performed} + \exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{owned\ regional}) + \varepsilon_{ict}$

$\Pr(VI = Outsourced\ to\ owned\ regional_{fatod} = \exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{owned\ regional} / (\exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{self-performed} + \exp\big(\beta_0 + \beta_1 * \widehat{D}elay_{ic} + \beta_2 * Distance_{od} + \beta_3 * Population_o + \beta_4 * Population_d + \beta_5 * Hub_o + \beta_6 * Hub_d + \beta_7 * Origin\ airport_o + \beta_8 * Destination\ airport_d + \beta_7 * Airline_a + \beta_8 * Time_t\big)_{owned\ regional}) + \varepsilon_{ict}$

A multinomial logit model is chosen in order to be able to test whether the effect of quality on vertical integration differs between vertical integration by performing a flight and vertical integration by outsourcing to an owned regional airline. The two parts of the MLM both estimate an effect of quality on vertical integration. As these coefficients are estimated in the same model, it is possible to test whether the coefficients differ significantly from each other. The MLM has one important assumption, namely the Independence of Irrelevant Alternatives (IIA) assumption (Kwak & Clayton-Matthews, 2002). This entails that the odds of one category versus another category does not depend on any other category available. This means that including or excluding any category should not affect the relative odds between the other categories. The IAA assumption can be tested by a Hausman diagnostic test, which drops one of the categories (Kwak & Clayton-Matthews, 2002).

Hypothesis 2 is tested based on the coefficients for the predicted values of the delay, β_1 in regression 3. It will be tested whether the two resulting coefficients are equal to each other or significantly differ from each other.

For hypothesis 2 to be supported, both coefficients have to be negative, similar to hypothesis 1. In addition, the coefficient in the model for vertical integration by outsourcing to an owned regional airline must be significantly lower than the coefficient in the model for vertical integration by a major airline performing a flight itself, as the effect of quality on the propensity to vertically integrate is hypothesised to be weaker for vertical integration by outsourcing to an owned regional airline compared to vertical integration by a major airline performing a flight itself. If this is true, hypothesis 2 is supported. Whether the coefficients are significantly different from each other, will be tested by a Wald test.

4. Results

This section will describe the results of the empirical analysis. First, the results regarding the first hypothesis are discussed. The first hypothesis expects a positive effect of quality on vertical integration. As quality is measured by delays, negative coefficients of quality are expected. After this, the results regarding the second hypothesis are discussed. The second hypothesis expects a weaker effect of quality on the propensity to vertically integrate by outsourcing to an owned regional airline compared to vertically integrate by a major airline by performing a flight itself. Next, the causality of the results and the validity of the instrumental variable approach are discussed. At last, it is discussed whether the hypotheses are supported or rejected.

4.1 Results for hypothesis 1

The results of the empirical analysis for the first hypothesis are presented in tables 4.1 and 4.2. Table 4.1 presents the first stage of the IV analysis and Table 4.2 the second stage. Average marginal effects for the predicted values of the arrival delays are presented in Table A.3 in the Appendix. The base model, model 1, will be discussed most extensively. Regarding models 2 to 5, only the changes will be discussed.

Table 4.1 Results of the first stage regression of precipitation on arrival delay

	Model				
Variable	(1)	(2)	(3)	(4)	(5)
Precipitation at origin airport (log)	-0.01***	-0.01***	-0.01***	0.05***	0.04***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Precipitation at destination airport (log)	-0.01***	-0.01***	-0.01***	0.07***	0.06***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Population at origin airport (log)	0.02***	0.03***	0.02***	0.27***	-0.07
	[0.00]	[0.00]	[0.00]	[0.05]	[0.05]
Population at destination airport (log)	0.04***	0.04***	0.04***	0.29***	-0.05
	[0.00]	[0.00]	[0.00]	[0.05]	[0.05]
Hub at origin airport (1 if true)	0.00	0.00	0.17***	-0.12***	0.07***
	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]
Hub at destination airport (1 if true)	0.00	0.00	0.18***	-0.15***	0.04***
	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]
Distance (log)	-0.03***	-0.03***	-0.03***	-0.04***	-0.04***
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Constant	1.71***	1.58***	1.54***	-4.12***	3.92***

	[0.02]	[0.02]	[0.02]	[0.91]	[0.85]
Observations	385,447	385,447	385,447	385,447	385,447
R^2	0.01	0.05	0.05	0.08	0.15
F-statistic	388.50	924.86	1038.21	1446.83	1548.23
Time Fixed Effects	No	Yes	No	No	Yes
Airline Fixed Effects	No	No	Yes	No	Yes
Airport Fixed Effects	No	No	No	Yes	Yes

Notes. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table 4.2 Results of the second stage regression of the predicted delay on the propensity to vertically integrate

	Model				
Variable	(1)	(2)	(3)	(4)	(5)
Predicted delay (log)	0.93***	0.82***	-0.61***	-0.39***	-1.28***
	[0.07]	[0.06]	[0.01]	[0.06]	[0.02]
Population at origin airport (log)	0.00	0.00	0.04***	-3.33***	2.63***
	[0.00]	[0.00]	[0.00]	[0.27]	[0.30]
Population at destination airport (log)	-0.01***	-0.01***	0.05***	-3.44***	2.53***
	[0.00]	[0.00]	[0.00]	[0.27]	[0.30]
Hub at origin airport (1 if true)	-0.74***	-0.76***	-0.73***	-1.96***	-2.12***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Hub at destination airport (1 if true)	-0.74***	-0.76***	-0.73***	-1.97***	-2.15***
	[0.01]	[0.01]	[0.01]	[0.02]	[0.02]
Distance (log)	0.86***	0.86***	0.82***	0.48***	0.46***
	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]
Constant	-6.81***	-6.34***	-4.18***	79.35***	-65.20***
	[0.12]	[0.11]	[0.04]	[4.13]	[5.22]
Observations	385,447	385,447	385,447	337,201	337,201
Time Fixed Effects	No	Yes	No	No	Yes
Airline Fixed Effects	No	No	Yes	No	Yes
Airport Fixed Effects	No	No	No	Yes	Yes

Notes. Robust standard errors in brackets. In models 4 and 5 48,246 observations are omitted due to perfect predictions of the outcome for the fixed effects of certain airports. *** p<0.01, ** p<0.05, * p<0.1

Model 1 controls for all control variables, except the fixed effects. The instrumental variable consists of the two precipitation variables. One variable measures the average precipitation at the origin airport, where the other variable measures the average precipitation at the destination airport. The coefficients of both variables are negative and significant at an 1% significance level. This means that if the average precipitation increases, the delay decreases. Explicitly, if the log of the average daily precipitation at the origin airport or the destination airport increases by 1%, the log of the arrival delay decreases by 0.01%, ceteris paribus. This is an interesting result as it was expected that more precipitation would make a flight more difficult and therefore result in greater delays. Contrary, model 1 predicts that delays would become smaller when precipitation increases.

The population variables are both positive and significant. This entails that if either endpoint has a higher population, delays increase, ceteris paribus. The coefficients of the hub variables are insignificant. Therefore, we cannot conclude anything about a hub effect. The coefficient of the last control variable, the distance variable, is negative and significant. So, if the distance between both endpoints increases, the delay decreases, ceteris paribus. This can be explained by the fact that if a flight is longer, there is more time to solve any delay that has occurred in the beginning of the flight or even before the start. In addition, when an airplane is flying at a constant speed and in the same direction for a long time, there aren't many reasons for a delay to occur.

The second stage of model 1 is presented in Table 4.2. A dummy indicating whether a flight has been performed by a major airline or an owned regional airline, when VI takes 1 as value, or by an independent regional airline, when VI takes 0 as value, is regressed at the predicted values of the delays by the first stage. The coefficient of the predicted values is positive and significant. This means that if delays are higher and quality thus lower, vertical integration is more likely. This is unexpected and contrary to the first hypothesis. The average marginal effect of this coefficient is 0.27, meaning that, on average, an 1% increase of the predicted delay leads to an increase in the probability of the flight being integrated by a major airline by 27 percentage points, ceteris paribus.

The population variables differ between origin airport and destination airport. Both coefficients are significant. The coefficient for the origin airport is positive, but the coefficient for the destination airport is negative. The destination airport coefficient is unexpected and contrary to the reasoning from Section 3.1.4 that larger cities create more demand and consequently the need for larger aircrafts. Only major airlines use these larger aircrafts. The distance variable is positive and significant. This is in line with the expectations, as longer flights are better suited for the type of aircrafts of major airlines, and also in line with the results of Forbes and Lederman (2009). They find a positive coefficient on the probability that a major airline performs a flight itself, thus on the propensity to vertically integrate. The coefficients on the hub variables are both significant and negative. This means that vertical integration is less likely if one of the endpoints is a hub of the performing airline. This is not in line with the expectation from Section 3.1.4 that hubs create an incentive for airlines to integrate a flight due to the higher importance of this route.

Models 2, 3 and 4 add some fixed effects to the control variables in model 1 in order to control for more unobserved characteristics. Model 2 controls for time fixed effects. These time fixed effects are based on the month and the year in which the flight is performed. The first stage does not change when adding time fixed effects and the second stage only changes to a minimal extent. The significance and directions of all the coefficients remains the same, but the magnitude of the coefficients changes. The coefficient of the predicted values of the delay changes from 0.93 to 0.82. The average marginal effect for model 2 is 0.23. This shows that seasonality or other time effects are not the cause of the positive coefficient of the delays on vertical integration.

When controlling for airline fixed effects in the first stage of model 3 there are also no important changes in coefficients. In the second stage, the coefficient of the predicted values remains significant, but changes from a positive coefficient to a negative coefficient, compared to models 1 and 2. This means that if delays are higher and quality thus lower, vertical integration is less likely, which is as expected and in line with the first hypothesis. The average marginal effect of this coefficient is -0.17. This means that, on average, an 1% increase in the predicted delay leads to a decrease of the probability of the flight being integrated by a major airline by 17 percentage points, ceteris paribus. In addition, in the second stage of model 3 both variables on the population are positive, as expected from Section 3.1.4. This result suggests that the positive coefficients of models 1 and 2 are caused by unobserved characteristics of specific airlines and do not reflect the true effect of quality on vertical integration.

At last, model 4 controls for airport fixed effects. If an airport is characterised by difficult circumstances, such as a difficult landing situation or a lot of traffic, delays are more likely. In the first stage the coefficients of the precipitation variables turn positive and remain significant. This is as expected, as more precipitation would result in larger delays. The coefficients of the hub variables are negative, compared to positive coefficients in models 1 to 3. The second stage is similar to model 3, with a significant and negative coefficient of the predicted values for the delay.

In addition, the population variables change significantly in model 4 compared to models 1, 2 and 3. The coefficients are in the other models significant and negative. When controlling for airport fixed effects, the population variables are significant but positive. This means that when there is a larger population at the endpoints, the probability of a major airline integrating a flight is higher. This is as expected and in line with the reasoning from Section 3.1.4.

Why does the inclusion of airline dummies or airport dummies affect the effect of quality on vertical integration? When including these dummies, a negative effect of delays on the propensity to vertically integrate is found, meaning that quality positively affects this propensity, as predicted. As a positive coefficient was estimated previously, there is a positive bias when excluding these dummies, meaning that the effect of delay on the propensity to integrate is estimated too high. This means that there are characteristics of airlines and airports that influence quality and the propensity to integrate and are not captured by the control variables.

Based on these 4 different models we cannot draw clear conclusions on the effect of quality on vertical integration. When controlling for time fixed effects, a negative coefficient is estimated. However, when controlling for airline fixed effects or airport fixed effects, a positive effect is estimated. Therefore, a model with all the fixed effects is estimated. This is model 5, which is the preferred model as all dummies are included.

The first stage of model 5 is different from the first stages of the previous models. The coefficients of the precipitation variables are positive and significant, as expected. This is in line with model 4, but contrary to models 1, 2 and 3. The coefficients of the population variables are negative and significant. These coefficients were positive in all the previous four models. The negative coefficient entails that when the population at the endpoints is larger, delays are smaller. This can be explained by the fact that airports in larger cities can be more important for airlines due to their high demand and therefore gain extra attention from airlines to prevent and minimise delays. The coefficients of the hub variables are positive and significant, meaning that delays are larger when a hub is the origin or destination airport. This is only similar to model 3, which controls for airline fixed effects.

The second stage of this model is interesting, especially compared to the previous models. The coefficients of the predicted values for the delay have a significant and negative coefficient. The average marginal effect is -0.25, meaning that, on average, an 1% increase of the predicted delay leads to a decrease of the probability of the flight being integrated by a major airline by 25 percentage points, ceteris paribus. This is in line with hypothesis 1 and similar to models 3 and 4, which respectively control for airline fixed effects and airport fixed effects. The population variables have a positive and significant coefficient. This is similar to models 1, 2 and 3, but contrary to model 4.

This model suggests that airline and airport fixed effects influence the estimation to a large extent, as described previously in this section. Model 5 is the most extensive model and is mostly similar to models 3 and 4, which respectively control for airline and airport fixed effects. Therefore, we can conclude that the effect of quality on the propensity to vertically integrate is mostly sensitive to airline- and airport-specific characteristics.

4.2 Results for hypothesis 2

Hypothesis 2 investigates whether the effect of quality on vertical integration is different between vertical integration by a major airline by performing a flight itself and vertical integration by outsourcing to an owned regional airline. The empirical analysis uses a multinomial logit model to analyse this hypothesis. Outsourcing to an independent regional airline is used as the baseline, as this is not a form of vertical integration. The first stages remain the same as the first stages for the analysis of hypothesis 2. With the same predicted values, five multinomial logit models are estimated, in the same manner as for hypothesis 1.

The results of the base model and most extensive model are presented below in Table 4.3, whereas the results of the other models are presented in Table A.4 in the Appendix. Each multinomial logit model is split into two parts, for example models 6a and 6b. The a-model estimates the regression of vertical integration by outsourcing to an owned regional airline relative to outsourcing to an independent regional airline. The b-model estimates the regression of vertical integration by performing a flight relative to outsourcing to an independent regional airline. Marginal effects and the results of a Wald test of the base model and the most extensive model are presented in Table 4.4. Marginal effects and the results of a Wald test of the other models are presented in Table A.5 in the Appendix.

Table 4.3 Results of the second stage multinomial logit regression of the predicted delay on the propensity to vertically integrate

	Model			
Variables	(6a)	(6b)	(10a)	(10b)
Predicted delay (log)	12.20 ***	3.73 ***	-2.01 ***	-2.01 ***
	[0.20]	[0.12]	[0.06]	[0.04]
Population at origin airport (log)	-0.37 ***	-0.06 ***	-2.69 *	-0.21
	[0.01]	[0.00]	[1.47]	[0.53]
Population at destination airport (log)	-0.56 ***	-0.12 ***	-2.66	-0.20
	[0.01]	[0.01]	[5.62]	[0.61]
Hub at origin airport (1 if true)	-0.54 ***	-1.39 ***	-3.08 ***	-4.44 ***
	[0.02]	[0.01]	[0.06]	[0.04]
Hub at destination airport (1 if true)	-0.56 ***	-1.39 ***	-3.14 ***	-4.49 ***
	[0.02]	[0.01]	[0.06]	[0.04]
Distance (log)	0.26 ***	1.49 ***	0.01	0.62 ***
	[0.01]	[0.01]	[0.02]	[0.01]
Constant	-19.23 ***	-14.42 ***	68.40	5.13
	[0.36]	[0.22]	[75.19]	[10.14]
Observations	385,447	385,447	385,447	386,447
Time Fixed Effects	No	No	Yes	Yes
Airline Fixed Effects	No	No	Yes	Yes

Airport Fixed Effects	No	No	Yes	Yes

Notes. Robust standard errors in brackets. Each model consists of an a-model and a b-model. The a-model estimates the regression of vertical integration by outsourcing to an owned regional airline relative to outsourcing to an independent regional airline. The b-model estimates the regression of vertical integration by performing a flight relative to outsourcing to an independent regional airline. *** p<0.01, ** p<0.05, * p<0.1

Table 4.4 Marginal effects and results of a Wald test from the multinomial logistic models

	Model	
Statistical variable or category	(6)	(10)
Wald test		
Chi value	1778.29***	0.00
P-value	0.00	0.95
Marginal effects		
Outsourcing to independent regional airline	-0.82 ***	0.20***
	(0.02)	(0.00)
Outsourcing to owned regional airline	0.64***	-0.04***
	(0.01)	(0.00)
Performed by airline	0.18***	-0.15***
	(0.02)	(0.00)

Notes. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

The first multinomial logit model is estimated without the fixed effects. Model 6a estimates a positive coefficient of the predicted values for the delays, which is significant at a 1% significance level. This means that an increase in predicted delays increases the probability of vertical integration by outsourcing to an owned regional airline relative to outsourcing to an independent regional airline. This means that a major airline is more likely to outsource to an owned regional airline than to an independent regional airline if delays are higher. The coefficient in model 6b is also positive. So, an increase in predicted delays increases the probability of vertical integration by a major airline performing a flight itself relative to outsourcing to an independent regional airline.

When looking at the average marginal effects, we can conclude more about the effects of quality on the different forms of vertical integration and outsourcing. The average marginal effect for category 1, meaning outsourcing to an independent regional airline, is -0.82. This coefficient is significant at an 1% significance level. This means that, on average, an increase of 1% in predicted delay leads to a decrease of the probability of outsourcing to an independent regional airline by 82 percentage points, ceteris paribus. This would mean that a lower quality would decrease the probability of outsourcing to an independent regional airline, which is not in line with the second hypothesis. The average marginal effect for vertical integration by outsourcing to an owned regional airline is 0.64. This coefficient is also significant at an 1% significance level. This coefficient means that, on average, an increase of predicted delay by 1% increases the probability of outsourcing to an owned regional airline by 64 percentage points, ceteris paribus. This is also contrary to hypothesis 2, which expects a positive effect of quality on the propensity to vertically integrate, so a negative coefficient of delays on the propensity. The same is the case for vertical integration by a major airline performing a flight itself. The average marginal effect for this type of vertical integration is 0.18. This coefficient is significant at a 1% significance level. This coefficient means that, on average, an increase of 1% of the predicted delay leads to a decrease of the probability of outsourcing to an owned regional airline by 18 percentage points, ceteris paribus.

The marginal effects for vertical integration are both positive, contrary to hypothesis 2. In addition, the hypothesised difference in the coefficients is also not found. It was hypothesised that the effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to an owned regional airline compared to vertical integration by a major airline performing a flight itself. The coefficients for the predicted values of delays are significantly different from each other in model 6. A Wald test finds a p-value of 0.00 on the hypothesis that the difference between both coefficients is equal to zero. Therefore, we can conclude that the coefficients are significantly different from each other at an 1% significance level. However, the marginal effect for vertical integration by a major airline performing a flight itself is smaller and thus weaker than for vertical integration by outsourcing to an owned regional airline. This means that both the directions of the coefficients and the difference between the coefficients is not in line with hypothesis 2.

The changes in coefficients when adding fixed effects in models 7 to 9 follow the same pattern in the multinomial logit models as in the probit models of hypothesis 1. In the base model and when controlling for time fixed effects, positive coefficients of delays on the propensity to vertically integrate are found. When controlling for airline fixed effects the coefficients turn negative, at least for 1 category in the multinomial logit model. Then, when controlling for airport fixed effects an insignificant coefficient is found. This pattern is observed in both the probit models and in the multinomial logit model. Interestingly, the Wald test becomes less significant in model 9, but still significant at a 10% significance level. As previously for hypothesis 1 the preferred full model followed the model with the airport fixed effects, this might be a sign that the difference between the coefficients might not be significant.

As last model, model 10 includes all the fixed effects. In this model both coefficients for the predicted values of the arrival delays are negative and significant at a 1% significance level. This means that an increase in predicted delays decreases the probability of vertical integration by outsourcing to an owned regional airline and by a major airline performing a flight itself, relative to outsourcing to an independent regional airline. This result is in line with hypothesis 2, as a major airline is less likely to vertically integrate if delays are higher. So, if quality is higher, vertical integration is more likely. The directions of the average marginal effects are now also in line with hypothesis 2, as they are negative. However, the coefficients are not significantly different from each other at a 10% significance level. The p-value of the Wald test is 0.95. Therefore, we cannot conclude that the coefficients are different from each other. This means that the positive effect of quality on the propensity to vertically integrate is not weaker for vertical integration by outsourcing to an owned regional airline compared to vertical integration by a major airline performing a flight itself.

As described in Section 3.3, the multinomial logit model assumes the independence of irrelevant alternatives. This assumption can be tested by a Hausman test. However, because the results already are not in line with hypothesis 2, it is not needed to perform this test. Therefore, no further analyses of the IAA assumption is conducted.

4.3 Causality

For an instrumental variable approach to establish a causal effect, several assumptions have to be satisfied. The first assumption is that the instrument has a clear and strong causal effect on the variable of interest or the independent variable. This is also called a strong first stage. The second assumption is that the instrument is uncorrelated with the error term, which is also called independence. The last assumption is that the instrument has no direct effect on the outcome or the dependent variable. This assumption is called the exclusion restriction.

For the regressions, a strong first stage means that precipitation, representing weather conditions, has a clear and strong causal effect on flight delays, representing quality. This assumption is often verified or falsified by looking at the significance of the coefficient of the instrumental variable and at the F-statistic of the first stage. As a rule of thumb, if the F-statistic of the first stage is higher than 10, the assumption is satisfied (Baiocchi, Cheng & Small, 2014). The coefficients of the precipitation variables are in all the first stage models highly significant. As presented in Table 4.1 and Table 4.3, all the coefficients are significant at a 1% significance level. The F-statistic is in all the models well above 10, with model 5, as the model with most control variables, with a F-statistic of 1548.23. Therefore, we can conclude that there is a clear and strong causal effect of average daily precipitation, which is the instrument, on flight delays, which is the variable of interest.

The second assumption on independence means that precipitation is not correlated with other determinants of flight delays. As weather conditions are not determined by airlines and are fully random, this assumption is satisfied. Because airlines cannot influence average daily precipitation, weather conditions are a clear exogenous variable. Therefore, the second assumption is also satisfied.

The last assumption, the exclusion restriction, cannot be tested (Van Kippersluis & Rietveld, 2018). Precipitation must not affect whether a major airline choose to perform a flight itself or outsource the flight to an owned regional airline on one hand or outsource the flight to an independent regional airline on the other hand. This assumption is difficult to verify. Especially, since Forbes and Lederman (2009) have found that weather conditions affect the choice whether to outsource to an owned or independent regional airline. However, it is important to look at what type of weather conditions are really directly affecting vertical integration and why. The reasoning of Forbes and Lederman (2009) is based on adaptations in the flight schedule due to adverse weather conditions. When adverse weather conditions lead to delays, the flight schedule has to be changed for connecting flights. This can easier be done when a flight has been integrated, compared to when it is outsourced to an independent regional airline, as a major airline can more easily change the schedule of an owned regional airline.

Forbes and Lederman (2009) use precipitation and snowfall as measures of adverse weather conditions. They find a consistent and robust positive effect of snowfall on vertical integration. However, the effect of precipitation is not clear. They only find a significant coefficient in their baseline model, but when adding more control variables there is no effect of precipitation on vertical integration. Also, when looking at non-hub routes versus hub routes the coefficient changes between positive, negative and insignificant. This shows that a direct effect of precipitation on vertical integration is not causal and not robust. As Forbes and Lederman (2009) explain, snowfall is much more likely to cause adaptations in the flight schedule than precipitation. Precipitation is a much more common event than snowfall and usually not adverse, so this would not directly affect vertical integration. Therefore, we can say that the exclusion restriction would not be violated.

To support this argument an extra test is performed on the results for hypothesis 1. As adverse weather conditions might affect vertical integration, we can remove adverse precipitation from the dataset. To do so, all the observations in the fourth quartile are removed from the dataset. This means that all the observations with adverse precipitation, which might lead to adaptations in the flight schedule, are removed. Model 5 is reestimated on this restricted dataset. The results are presented in table A.6 in the Appendix. The coefficients do not change in direction or significance, compared to the models which are estimated on the unrestricted dataset. This shows that adverse precipitation does not affect the results. This approach excludes observations in which adaptations in the flight schedule might play a role, which means that a direct effect of precipitation on vertical integration is less likely to be present.

These arguments and the robustness test suggest that a direct effect of precipitation on the propensity to vertically integrate is not present, so that the exclusion restriction is not violated. However, we can never completely test this assumption and be completely sure about this. The robustness test only indicates this.

As explained, two assumptions of an instrumental variable model are satisfied and there are indications that the third assumption is likely to be satisfied. This means that there are signs that the positive effect of quality found in the empirical analysis is a causal effect, although this is not certain.

4.4 Robustness checks

In order to strengthen the results found in the previous sections, three robustness checks are performed. In the previous section a robustness test on the precipitation variable was already performed, which regarded one variable. To also test the robustness of the results on other aspects, one robustness test is performed with respect to the sample and two other robustness checks are performed with respect to the methods. For these robustness test only model 5 is reestimated. The results of the robustness checks are presented in Tables A.7 and A.8 in the Appendix.

The first robustness test to check the sample uses a sample without low-cost airline. This can show whether the obtained results are clear and not obscured with low-cost airlines that never outsource. As described in Section 2.3 low-cost airlines have different characteristics, which may affect the results. The results of this analysis shows similar results. In the first stage of model 11 the coefficients of precipitation are 0.05 and 0.06 and significant, compared to 0.04 and 0.06 in the preferred model 5. The second stage is similar as well with a significant negative coefficient. Therefore, we can conclude that the results are robust to the exclusion of low-cost airlines.

The second robustness test uses a Tobit model as first stage, which is model 12. As arrival delays cannot be lower than zero, this variable is censored at zero. However, airplanes can also arrive earlier, but are then observed with a delay of zero. For this kind of data a Tobit model can be estimated (Austin, Escobar & Kopec, 2000). The Tobit model estimates the first stage and a probit model is used for the second stage, as in the main methodology. This analysis also estimates similar results to model 5. This shows that the results are not sensitive to the delay variable being non-negative.

The third robustness test uses an OLS model instead of a probit model as second stage, which is model 13. The second stage of model 5 does not consist of all the observations, as some of the dummies for the airports predict the outcome perfectly. It might be possible that the omitted observations could change the results. An OLS regression allows for a dummy to predict the outcome perfectly and therefore can include all the observations. This analysis also estimates similar models. As an addition, the OLS model allows to estimate a R^2, to show what proportion of the variance in the propensity to vertically integrate is explained by the model. The R^2 is 0.51. This means that 51% of the variance is explained, which is a high proportion. The results of this robustness test show that the methodology does not affects the results.

As explained here, the robustness checks support the results found previously and do not rise concerns.

4.5 Support for or rejection of the hypotheses

This section discusses whether the hypotheses are supported or rejected. The first hypothesis states that quality has a positive causal effect on the propensity to vertically integrate. This hypothesis is supported, based on model 5. When controlling for all control variables and all fixed effects, a positive effect of quality on the propensity to vertically integrate is found with an instrumental variable method. As discussed, there are signs that this effect is a causal effect, which means that the data support hypothesis 1.

Hypothesis 2 differentiates between vertical integration by a major airline performing a flight itself and vertical integration by outsourcing to an owned regional airline. The hypothesis states that the positive causal effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to an owned regional airline compared to vertical integration performing a flight. The empirical analysis does not support this hypothesis, which means that hypothesis 2 is rejected. In the most extensive instrumental variable model with a multinomial logit model as second stage, model 10, the coefficients of the predicted values of the arrival delay of the two categories of vertical integration do not significantly differ from each other. Therefore, we cannot conclude that the positive causal effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to an owned regional airline compared to vertical integration by performing a flight.

5. Discussion & Conclusion

In this study the following research question has been studied: does quality affect vertical integration in the airline industry? This section discusses the answer to this research question. First, the results are described. Then, the results are discussed based on validity and limitations, whereafter the results are connected to the economic literature. Following this discussion, a conclusion is drawn and the research question is answered. At last, suggestions for further research are presented and the implications for theory and practice are discussed.

5.1 Overview of the results

The main result of this research is that there is a positive relationship between quality and the propensity to vertically integrate. This effect is most likely a causal effect. This result has been found by an instrumental variable analysis. In this analysis, quality, measured by arrival delays, was regressed on weather conditions, measured by precipitation. The regression controlled for the population of the endpoints, the presence of a hub at an endpoint and the distance between both endpoints. In addition, fixed effects were included to control for time effects, airline characteristics and airports characteristics. Then, using a probit regression, a dummy indicating whether the flight had been vertically integrated was regressed on the predicted values of the arrival delays. This regression controlled for the same variables and fixed effects.

A robustness test showed that the exclusion of observations with adverse precipitation had no influence on the direction and significance of the effect that was found in the analysis. This gives a good indication that precipitation has no effect on vertical integration, in light of the results of Forbes and Lederman (2009) on an effect of adverse weather conditions on vertical integration. Other robustness tests also presented no issues. Therefore, the first hypothesis stating that quality has a positive causal effect on the propensity to vertically integrate, was supported.

The second hypothesis investigated the difference between vertical integration by performing a flight and vertical integration by outsourcing to an owned regional airline. The hypothesis stated that the positive casual effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to owned regional airlines compared to vertical integration by performing flights.

This hypothesis was investigated by an instrumental variable analysis with a multinomial logit model as second stage. In the multinomial logit model that included all of the fixed effects, the coefficients for the predicted values of the arrival delay did not significantly differ between the model estimating vertical integration by performing a flight relative to outsourcing to an independent regional airline and the model estimating vertical integration by outsourcing to an owned regional airline relative to outsourcing to an independent regional airline. This means that there is no support that the effect of quality on the propensity to vertically integrate is weaker for vertical integration by outsourcing to owned regional airlines compared to vertical integration by performing a flight. Therefore, hypothesis 2 was rejected.

We can conclude that there is a positive relationship between quality and vertical integration, but that there is no difference between integration by performing a flight and integration by outsourcing to owned regional airlines. There are signs that the relationship is a causal effect.

5.2 Discussion of the results

This section discusses the limitations of this study, divided between internal and external validity. After this, the results will be connected to the economic literature that has been discussed in the theoretical framework.

5.2.1 Internal validity

The main concern and limitation of this study regards the causality of the effect of quality on the propensity to vertically integrate, as discussed in Section 4.3. An instrumental variable model has three assumptions, namely a strong first stage, independence of the instrument and the exclusion restriction. A strong first stage and independence of the instrument were confirmed. However, whether the exclusion restriction is satisfied remains uncertain. This assumption cannot be formally tested and often leads to uncertainty around results when using an instrumental variable approach (Van Kippersluis & Rietveld, 2018).

Usually, weather conditions are clear to be exogenous and therefore serve well as an instrumental variable. However, in the case of vertical integration the study of Forbes and Lederman (2009) shows that might not be true. They find a direct effect of adverse weather conditions on vertical integration. This suggests that the exclusion restriction is violated. However, when looking closely at the results of Forbes and Lederman (2009) this is not the case, as precipitation has not a clear and robust effect on vertical integration. Only adverse weather conditions, as snowfall, have a clear and robust effect on vertical integration.

To further investigate this limitation, the baseline model and the most extensive model were reestimated on a restricted sample. In this sample all the observations in the fourth quartile were excluded. This means that there were no cases with adverse precipitation in the dataset. The results in the empirical analysis were robust to this test. As adverse weather conditions were no longer present in the dataset, it is suggested that precipitation would not have a direct effect on vertical integration. Therefore, there are signs that the positive effect of quality on the propensity to vertically integrate that was found, is a causal effect. However, we can never formally test this, so can never be completely certain about this. We can be certain that there is a positive relationship between quality and the propensity to vertically integrate, after controlling for population, distance, the presence of a hub, time effects, airport characteristics and airline characteristics. There are signs that this is also a causal effect.

Another potential limitation is that the data on arrival delays per flight are averaged per month. This was done to make the dataset better workable, but reduces the number of observations and might reduce the variation in the dataset. However, this loss of variation most likely will not affect the results. Vertical integration does not change per flight, so the specific delay of a specific flight at a specific time most likely does not affect the propensity to vertically integrate. It is more likely that this propensity changes with a structural low quality, represented by structural delays. Therefore, this limitation has no consequences for the results.

Two other choices in the data selection process could impose a limitation. First, it was decided to only incorporate direct flights. Therefore, common routes with a transfer are not recognised and indirect flights are recognised as two or more direct flights. This means that any effects of transfers and connecting flights could not be analysed. However, it is not expected that this would have changed the principal result of this study that there is a positive relationship between quality and vertical integration. It could only have been an interesting moderation of this relationship. The second choice in the data selection was to only include airlines that account for at least 1% of the domestic passenger revenue in the US, which means that very small airlines are not included. This choice was made by the U.S. Bureau of Transport Statistics, as the data were retrieved from a dataset of the Bureau. This limitation is considered not to have affected the results, as the data were distributed among five major airlines, seven regional airlines and six low-cost airlines. This means that there was enough data on all types of airlines, such that excluded airlines would not have affected the results. This is also true, as the excluded airlines would only have a small number of flights, which would be irrelevant regarding the high number of observations.

At last, there can also be limitations regarding the variables. The variables on arrival delays, precipitation, distance and population are skewed. Therefore, the logarithmic values of these variables were used. After this procedure, the distributions were only skewed to a small extent, such that it is expected that this limitation does not affect the results. Another limitation is that the instrumental variable weather conditions was only determined by precipitation. Other relevant weather conditions, such as snowfall, wind or fog were not used. This could affect the results, as there might be airports with not much precipitation, but which often experience snowfall, wind or fog. However, it is expected that this would only affect the results to a minimum extent, as usually places with these weather conditions also experience precipitation. There most likely are only a limited number of airports in the dataset that experience not much precipitation but do experience often other weather conditions. This limited number would not affect the results regarding the high number of observations. In addition, severe weather conditions, such as snowfall and fog, could also affect the validity of the results in a negatively, as there might be a direct effect of adverse weather conditions on vertical integration, relating to the results of Forbes and Lederman (2009). Measurement errors most likely are not a limitation. The variables on arrival delays, precipitation, population and distance are retrieved from datasets from the U.S. government that are almost complete. Therefore, there are no missing values and most likely also not much measurement errors.

5.2.2 External validity

This study finds a positive relationship between quality and the propensity to vertically integrate in the airline industry, which might be a causal effect. The airline industry serves as a good environment to study quality and vertical integration, because quality is well observable by on-time performance, measured by delays, and because of the difference in characteristics between major airlines and regional airlines that can results in incentives for both vertical integration and outsourcing, such as the difference in experience of employees.

However, this brings up the question whether the results are also valid for other industries. Does quality also affect the propensity to vertically integrate in other industries? This question is difficult to answer, as the airline industry is a specific industry with specific characteristics. As discussed in the theoretical framework, research has already found that vertical integration is more likely when quality is an important factor for customers (Lin et al., 2014). This could be extended to the results of this study, meaning that the positive relationship between quality and the propensity to vertically integrate might only be true for industries in which quality is an important factor for customers, such as the wine industry and the health care industry (Fernández-Olmos et al., 2019; Coles & Hesterly, 1998).

Another limitation with regards to the external validity is that the U.S. airline industry has been studied. This means that the results might not be valid for other countries. For example, even the European airline industry is not characterised by major airlines and outsourcing to regional airlines, but by joined ventures and outsourcing to owned low-cost airlines. This could mean that there might be no relationship between quality and vertical integration in the European airline industry. This means that the positive relationship between quality and the propensity to vertically integrate might not be directly applicable to the other airline industries.

5.2.3 Link to the economic literature

In the theoretical framework, it was described that not many economic researches had yet studied the effect of quality on vertical integration. Most of the literature had focused on transaction costs, such as Forbes and Lederman (2009) for the airline industry. Only Fernández-Olmos et al. (2019) and Coles and Hesterly (1998) had studied this relationship, respectively in the wine industry and the health care industry. Both studies had found a positive effect of quality on vertical integration, due to the risk of reputation loss. The result of the current study that quality is also positively related to the propensity to vertically integrate in the airline industry, so is in line with the studies of Fernández-Olmos et al. (2019) and Coles and Hesterly (1998). This result shows that the risk of reputation is an important factor for firms when deciding whether to vertically integrate. Outsourcing to independent firms without control over the quality or without the ability to measure the quality that is set by the independent firm, can create a risk for the reputation of firms. For firms with a reputation of high quality, this risk can outweigh the (cost) benefits of outsourcing. This reasoning has been supported by the results of this study, as a positive effect of quality on vertical integration was found.

5.3 Conclusion

This research investigated the following research question: does quality affect vertical integration in the airline industry? Based on the results of the empirical analysis in which an instrumental variable method was used, it can be concluded that there is a positive relationship between quality and the propensity to vertically integrate in the airline industry. This effect is most likely a causal effect, but certainly a relationship which is robust to controlling for many characteristics. There is no evidence that the effect of quality on the propensity to vertically integrate is different between vertical integration by a major airline performing a flight itself and vertical integration by outsourcing to an owned regional airline.

5.3.1 Suggestions for further research

As discussed, only two previous studies also have found that quality affects vertical integration. As the current study confirms this, research could further develop the understanding of this relationship. So far, the studies have been set in industries in which customers find quality important. This is the case for the wine industry, health care industry and, as discussed in Section 2.4, also for the airline industry. Therefore, it would be interesting to study whether the relationship between quality and vertical integration is also present in industries in which quality does not play a large role. Such a relationship is not expected based on the theoretical and mathematical findings of Lin et al. (2014) that vertical integration is more likely in industries in which consumers are more sensitive to quality. However, for a thorough understanding of the relationship between quality and vertical integration it would be good if empirical studies could confirm or contradict the findings of Lin et al. (2014). For example, the automotive industry can be studied. Monteverde and Teece (1982) have studied the effects of transaction costs on vertical integration in this industry, but as reputation is very important for automotive manufacturers, quality could also have an effect.

In addition, it would be useful for further research to study any moderators or mediators of the relationship between quality and vertical integration. So far, none of the discussed studies has done this, as studying the sole effect of quality on vertical integration was already new in the economic literature. As the current study has also found a relationship between quality and vertical integration, it is important to increase the understanding of this effect by investigating potential moderators and mediators.

5.3.2 Implications for economic theory and societal practice

The implications for economic theory are connected to the suggestions for further research. As discussed, an effect of quality on vertical integration had already been found by two studies, which this study supports. This suggests that quality is a variable that more often should be incorporated into economic theory and economic reasoning, as quality can be a strategic variable for firms, as argued in Section 2.4.

For society the results imply that independent regional airlines should improve their quality in order to make it attractive for major airlines to outsource more flights to these airlines instead of integrating these flights. For government, this study has no direct implications on how to improve quality in the airline industry. However, this study has shed new light on the interplay between quality and vertical integration, which can help in further discussions on how to improve quality in the airline industry.

Bibliography

Alaska Airlines. (2020). *Network*. Retrieved from https://newsroom.alaskaair.com/network

American Airlines. (n.d.). *Flagship lounge*. Retrieved from https://www.aa.com/i18n/travel-info/clubs/flagship-lounge.jsp

Arocena, P. (2008). Cost and quality gains from diversification and vertical integration in the electricity industry: A DEA approach. *Energy economics*, *30*(1), 39-58.

Austin, P.C. Escobar, M. & Kopec, J.A. (2000). The use of the Tobit model for analyzing measures of health status. *Quality of Life Research*, *9*(8), 901-910.

Baiocchi, M., Cheng, J. & Small, D.S. (2014). Instrumental variable methods for causal inference. *Statistics in medicine*, *33*(13), 2297-2340.

Blair, M.M. (2003). The economics of post-September 11 financial aid to airlines. *Indiana Law Review, 36*(2), 367-396.

Borenstein, S. (1992). The evolution of US airline competition. *Journal of Economic perspectives*, *6*(2), 45-73.

Brueckner, J.K., Lee, D. & Singer, E.S. (2013). Airline competition and domestic US airfares: A comprehensive reappraisal. *Economics of Transportation*, *2*(1), 1-17.

Bureau of Transportation Statistics. (n.d.). *On-time: Reporting carrier on-time performance (1987-present).* Retrieved from https://www.transtats.bts.gov/DL_SelectFields.asp?gnoyr_VQ=FGJ&QO_fu146_anzr=b0-gvzr

Carlton, D.W. (1979). Vertical integration in competitive markets under uncertainty. *The Journal of Industrial Economics, 27*(3), 189-209.

Cento, A. (2009). *The airline industry: Challenges in the 21st century.* Heidelberg: Physica-Verlag.

Coles, J.W. & Hesterly, W.S. (1998). Transaction costs, quality, and economies of scale: examining contracting choices in the hospital industry. *Journal of Corporate Finance*, *4*(4), 321-345.

Curto, J.D. & Pinto, J.C. (2011). The corrected vif (cvif). *Journal of Applied Statistics*, *38*(7), 1499-1507.

D'Aveni, R.A. & Ravenscraft, D.J. (1994). Economies of integration versus bureaucracy costs: does vertical integration improve performance?. *Academy of management Journal*, *37*(5), 1167-1206.

Delta. (2021). *Corporate stats and facts*. Retrieved from https://news.delta.com/corporate-stats-and-facts

FAA. (2021). *Fact sheet – inclement weather*. Retrieved from https://www.faa.gov/news/fact_sheets/news_story.cfm?newsId=23074#:~:text=Inclement%20weather%2C%20including%20thunderstorms%2C%20snowstorms,largest%20cause%20of%20flight%20delays.

Fernández-Olmos, M., Rosell-Martínez, J. & Espitia-Escuer, M. A. (2009). Vertical integration in the wine industry: a transaction costs analysis on the Rioja DOCa. *Agribusiness: An International Journal*, *25*(2), 231-250.

Forbes, S.J. & Lederman, M. (2007). The role of regional airlines in the US airline industry. *Advances in Airline Economics, 2*, 193-208.

Forbes, S.J. & Lederman, M. (2009). Adaptation and vertical integration in the airline industry. *American Economic Review*, *99*(5), 1831-49.

Forbes, S.J. & Lederman, M. (2010). Does vertical integration affect firm performance? Evidence from the airline industry. *The RAND Journal of Economics*, *41*(4), 765-790.

Goetz, A.R. & Vowles, T.M. (2009). The good, the bad, and the ugly: 30 years of US airline deregulation. *Journal of Transport Geography*, *17*(4), 251-263.

Grossman, S.J. & Hart, O.D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, *94*(4), 691-719.

Harrigan, K.R. (1984). Formulating vertical integration strategies. *Academy of management review*, *9*(4), 638-652.

Hawaiian Airlines. (n.d.). *Where we fly.* Retrieved from https://www.hawaiianairlines.com/destinations

Hunter, L. (2006). Low cost airlines: Business model and employment relations. *European Management Journal*, *24*(5), 315-321.

Kippersluis, H. van & Rietveld, C.A. (2018). Beyond plausibly exogenous. *The Econometrics Journal*, *21*(3), 316-331.

Kwak, C. & Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing research*, *51*(6), 404-410.

Lafontaine, F. & Slade, M. (2007). Vertical integration and firm boundaries: The evidence. *Journal of Economic Literature, 45*(3), 629-685.

Lin, Y.T., Parlaktürk, A.K. & Swaminathan, J.M. (2014). Vertical integration under competition: Forward, backward, or no integration? *Production and Operations Management*, *23*(1), 19-35

Maddigan, R.J. (1981). The measurement of vertical integration. *The Review of Economics and Statistics, 63*(3), 328-335.

Mahoney, J.T. (1992). The choice of organizational form: vertical financial ownership versus other methods of vertical integration. *Strategic Management Journal*, *13*(8), 559-584.

Mazzeo, M.J. (2003). Competition and service quality in the US airline industry. *Review of industrial Organization*, *22*(4), 275-296.

Monteverde, K. & Teece, D.J. (1982). Supplier switching costs and vertical integration in the automobile industry. *The Bell Journal of Economics, 13*(1), 206-213.

NOAA. (n.d.). *Weather and climate resources.* Retrieved from https://www.noaa.gov/weather-and-climate-resources

Ostrowski, P.L., O'Brien, T.V. & Gordon, G.L. (1993). Service quality and customer loyalty in the commercial airline industry. *Journal of travel research*, *32*(2), 16-24.

RAA. (n.d.). *RAA annual reports*. Retrieved from https://www.raa.org/content-hub/raa-annual-reports/

Reeven, P. van & Pennings, E. (2016). On the relation between multimarket contact and service quality: mutual forbearance or network coordination? *Strategic Management Journal, 37*(10), 2121-2134.

Short, M.N. & Ho, V. (2020). Weighing the effects of vertical integration versus market concentration on hospital quality. *Medical Care Research and Review, 77*(6), 538-548.

Suzuki, Y. (2000). The Relationship between on-time performance and airline market share: A new approach. *Transportation Research Part E, 36*(2), 139–154.

T4. (2021). *US airline market share.* Retrieved from https://www.t4.ai/industry/us-airline-market-share

United States Census Bureau. (n.d.). *City and town population totals: 2010-2019.* Retrieved from https://www.census.gov/data/datasets/time-series/demo/popest/2010s-total-cities-and-towns.html

United. (n.d.). *Airports and terminal maps*. Retrieved from https://www.united.com/ual/en/us/fly/travel/airport/maps.html

Whinston, M.D. (2001). Assessing the property rights and transaction-cost theories of firm scope. *American Economic Review*, *91*(2), 184-188.

Williamson, O.E. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law & Economics, 22*(2), 233-262.

Zeithaml, V.A., Berry, L.L. & Parasuraman, A. (1996). The behavioral consequences of service quality. *Journal of marketing*, *60*(2), 31-46.

Appendix

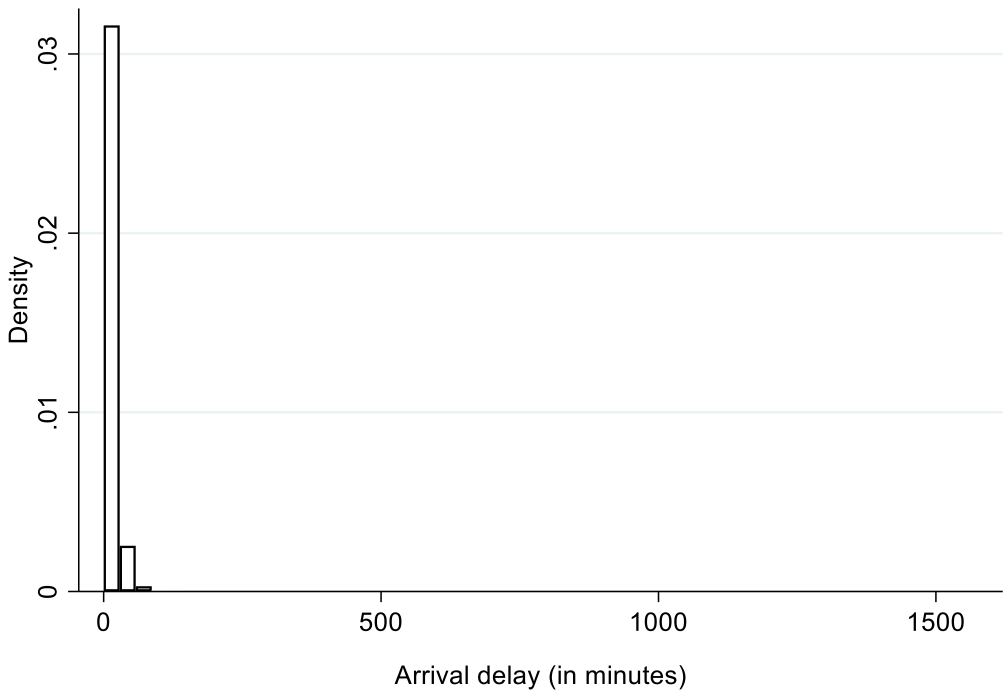

Figure A.1 Histogram of the variable on the arrival delays

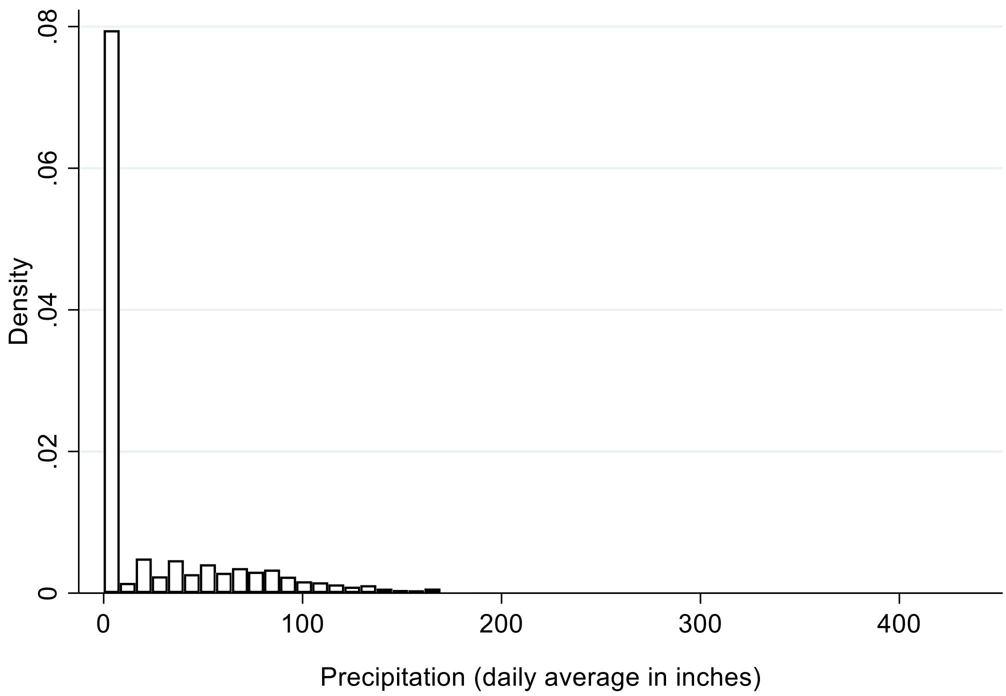

Figure A.2 Histogram of the variable on precipitation at airports

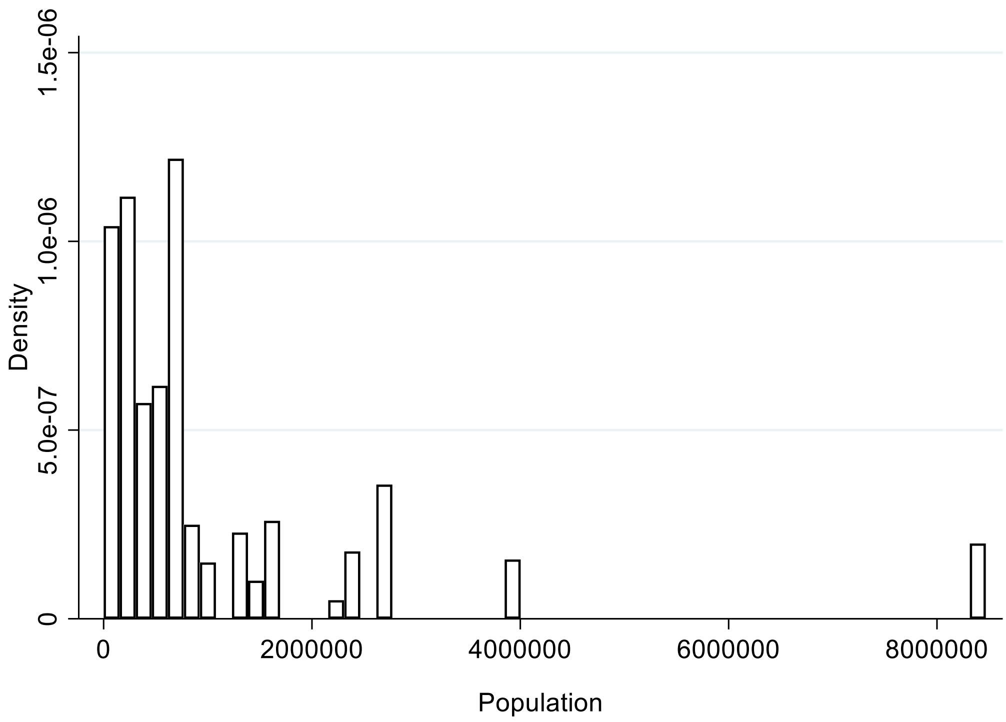

Figure A.3 Histogram of the variable on the population in origin cities

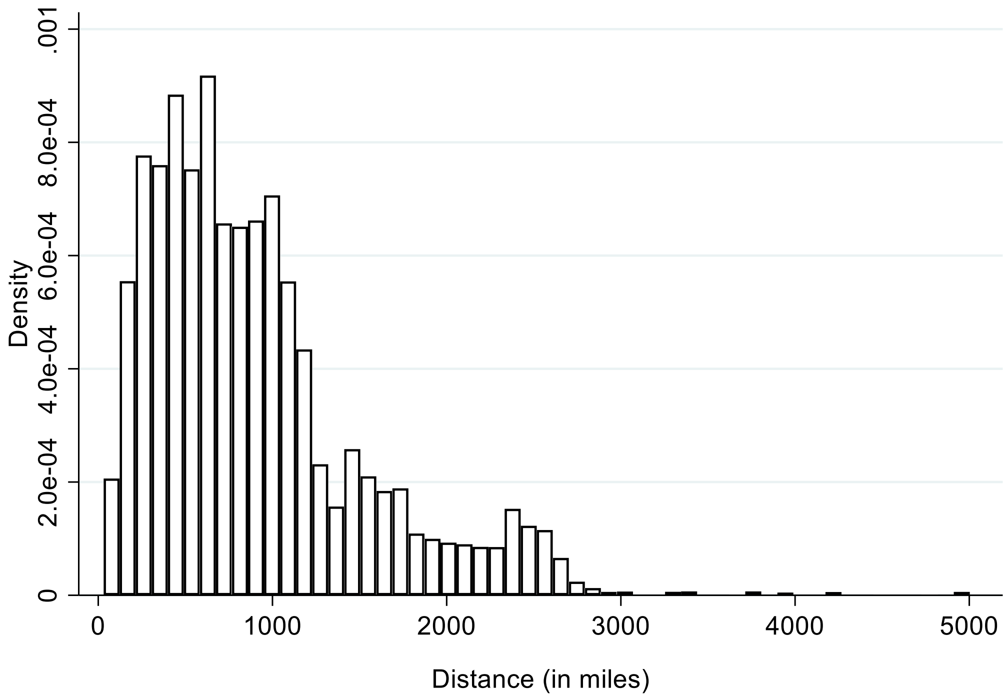

Figure A.4 Histogram of the variable on the distance between the endpoints of a flight

Table A.1 Regression results of the VIF test

	Model	
	(1)	(2)
Variable	On delay	On VI
Precipitation at origin airport (log)	-0.01***	
	[0.00]	
Precipitation at destination airport (log)	-0.01***	
	[0.00]	
Arrival delay (log)		-0.01***
		[0.00]
Population at origin airport (log)	0.02***	0.01***
	[0.00]	[0.00]
Population at destination airport (log)	0.04***	0.01***
	[0.00]	[0.00]
Hub at origin airport (1 if true)	0.00	-0.23***
	[0.00]	[0.00]
Hub at destination airport (1 if true)	0.00	-0.23***
	[0.00]	[0.00]
Distance (log)	-0.03***	0.25***
	[0.00]	[0.00]
Constant	1.71***	-1.08***
	[0.02]	[0.01]
Observations	385,447	385,447
R^2	0.01	0.25

Notes. Robust standard errors in brackets. Both models have been estimated with an ordinary least squares regression without any fixed effects dummies. *** p<0.01, ** p<0.05, * p<0.1

Table A.2 VIF scores of the regression in table A.1

	Model	
	(1)	(2)
Variable	First stage	Second stage
Precipitation at origin airport (log)	1.03	
Precipitation at destination airport (log)	1.03	
Arrival delay (log)		1.01
Population at origin airport (log)	1.18	1.19
Population at destination airport (log)	1.18	1.18
Hub at origin airport (1 if true)	1.15	1.15
Hub at destination airport (1 if true)	1.15	1.15
Distance (log)	1.05	1.05

Table A.3 Average marginal effects for models 1 to 5.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Predicted	0.27***	0.23 ***	-0.17***	-0.08	-0.25***
arrival delay	[0.02]	[0.02]	[0.00]	[0.01]	[0.00]

Notes. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A.4 Results of the second stage with a multinomial logit model for the models with only one type of fixed effects

	Model					
Variables	(7a)	(7b)	(8a)	(8b)	(9a)	(9b)
Predicted delay (log)	12.11 ***	3.37 ***	2.41 ***	-0.59 ***	0.28	-0.26 **
	[0.19]	[0.11]	[0.03]	[0.02]	[0.32]	[0.11]
Population at origin airport (log)	-0.39 ***	-0.06 ***	-0.14 ***	0.04 ***	3.03 **	0.16
	[0.01]	[0.00]	[0.00]	[0.00]	[1.39]	[0.53]
Population at destination airport	-0.58 ***	-0.12 ***	-0.18 ***	0.05 ***	3.01 ***	0.14
(log)	[0.01]	[0.01]	[0.00]	[0.00]	[0.68]	[0.48]
Hub at origin airport (1 if true)	-0.46 ***	-1.41 ***	-0.51 ***	-1.39 ***	-2.20 ***	-3.83 ***
	[0.02]	[0.01]	[0.02]	[0.01]	[0.07]	[0.03]
Hub at destination airport (1 if	-0.48 ***	-1.41 ***	-0.52 ***	-1.39 ***	-2.21 ***	-3.83 ***
true)	[0.02]	[0.01]	[0.02]	[0.01]	[0.07]	[0.03]
Distance (log)	0.15 ***	1.49 ***	-0.07 ***	1.39 ***	0.13 ***	0.62 ***
	[0.01]	[0.01]	[0.01]	[0.01]	[0.02]	[0.01]

Constant	-37.23	-13.22 ***	-2.18 ***	-7.03 ***	-75.44 ***	-7.28
	[0.00]	[0.19]	[0.08]	[0.08]	[18.20]	[7.74]
Observations	385,447	385,447	385,447	385,447	385,447	385,447
Time Fixed Effects	Yes	Yes	No	No	No	No
Airline Fixed Effects	No	No	Yes	Yes	No	No
Airport Fixed Effects	No	No	No	No	Yes	Yes

Notes. Robust standard errors in brackets. Each model consists of an a-model and a b-model. The a-model estimates the regression of vertical integration by outsourcing to an owned regional airline relative to outsourcing to an independent regional airline. The b-model estimates the regression of vertical integration by performing a flight relative to outsourcing to an independent regional airline. *** p<0.01, ** p<0.05, * p<0.1

Table A.5 Marginal effects and results of a Wald test from the multinomial logistic models with only one type of fixed effects presented in table A.4

	Model		
	(7)	(8)	(9)
Wald test			
Chi value	2042.57***	8478.56***	2.74*
P-value	0.00	0.00	0.10
Marginal effects			
Outsourcing to independent	-0.75***	0.01***	0.01
regional airline	[0.02]	[0.00]	[0.01]
Outsourcing to owned regional	0.59***	0.17***	0.02
airline	[0.01]	[0.00]	[0.02]
Performed by airline	0.16***	-0.18***	-0.03**
	[0.02]	[0.00]	[0.01]

Notes. Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A.6 Results of the robustness check on causality

Variables	First stage of model 5 on arrival delay	Second stage of model 5 on VI
Adjusted precipitation at destination airport (log)	0.04***	
	[0.00]	
Adjusted population at origin airport (log)	0.06***	
	[0.00]	

Predicted delay (log)		-1.13***
		[0.03]
Population at origin airport (log)	-0.15**	0.38
	[0.07]	[0.38]
Population at destination airport (log)	-0.09	0.41
	[0.05]	[0.38]
Hub at origin airport (1 if true)	0.06***	-1.91***
	[0.01]	[0.02]
Hub at destination airport (1 if true)	0.05***	-1.91***
	[0.01]	[0.02]
Distance (log)	-0.03***	0.51***
	[0.00]	[0.01]
Constant	5.27***	-12.06*
	[1.07]	[6.91]
Observations	220,371	197,989
R^2	0.154	
Time Fixed Effects	Yes	Yes
Airline Fixed Effects	Yes	Yes
Airport Fixed Effects	Yes	Yes

Notes. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A.7 Results on the first stage of the robustness checks

	Model		
Variable	(11)	(12)	(13)
	OLS on restricted sample	Tobit on full sample	OLS on full sample
Precipitation at origin airport (log)	0.05***	0.04***	0.04***
	[0.00]	[0.00]	[0.00]
Precipitation at destination airport (log)	0.06***	0.06***	0.06***
	[0.00]	[0.00]	[0.00]
Population at origin airport (log)	-0.08	-0.07	-0.07
	[0.08]	[0.05]	[0.05]
	-0.11	-0.06	-0.05

Population at destination airport (log)	[0.08]	[0.05]	[0.05]
Hub at origin airport (1 if true)	0.07***	0.08***	0.07***
	[0.01]	[0.01]	[0.01]
Hub at destination airport (1 if true)	0.02**	0.04***	0.04***
	[0.01]	[0.01]	[0.01]
Distance (log)	-0.01***	-0.04***	-0.04***
	[0.00]	[0.00]	[0.00]
var(e.Delay)		0.72***	
		[0.00]	
Constant	4.56***	3.94***	3.92***
	[1.40]	[0.85]	[0.85]
Observations	249,706	385,447	385,447
R^2	0.141		0.145
Time Fixed Effects	Yes	Yes	Yes
Airline Fixed Effects	Yes	Yes	Yes
Airport Fixed Effects	Yes	Yes	Yes

Notes. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1

Table A.8 Results on the second stage of the robustness checks

	Model		
	(11)	(12)	(13)
Variable	Probit on restricted sample	Probit after Tobit	OLS
Predicted delay	-2.30***	-1.03***	-0.24***
	[0.02]	[0.02]	[0.00]
Population at origin airport (log)	3.19***	2.61***	0.06***
	[0.33]	[0.30]	[0.01]
Population at destination airport (log)	2.95***	2.50***	0.06***
	[0.33]	[0.30]	[0.01]
Hub at origin airport (1 if true)	-1.06***	-2.08***	-0.30***
	[0.02]	[0.02]	[0.00]
Hub at destination airport (1 if true)	-1.16***	-2.11***	-0.32***
	[0.02]	[0.02]	[0.00]

Distance (log)	0.40***	0.46***	0.10***
	[0.01]	[0.01]	[0.00]
Constant	-76.08***	-65.20***	-1.39***
	[5.77]	[5.22]	[0.20]
Observations	213,040	337,201	385,447
R^2			0.505
Time Fixed Effects	Yes	Yes	Yes
Airline Fixed Effects	Yes	Yes	Yes
Airport Fixed Effects	Yes	Yes	Yes

Notes. Robust standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1