What Does Governance Do? Evidence From Airlines*

Silke Forbes† Ricard Gil‡ Giorgio Zanarone§

September 7, 2026

Abstract

Theories of the firm have identified asset ownership and employee management as two distinct reasons why vertical integration may outperform outsourcing contracts. Yet, while the empirical literature has found support for the ownership channel, the employment channel remains understudied. In this paper, we provide evidence that vertical integration increases performance even after controlling for asset ownership effects. Using rich data from the US airline industry, we show that flights ticketed by a major airline are less delayed if operated by the major's employees than by the employees of an outsourced regional partner, holding constant (a) the major's ability and (b) the major's incentive to allocate relevant assets across the two types of flights. This persistent, ownership-independent effect of vertical integration on performance is consistent with the view that firms more effectively manage employees within boundaries than across, thanks to the unique contracting rights created by employment relationships within firms (Williamson, 1991; Holmstrom, 1999).

JEL Codes: D22, D23, L14, L24, L93

*We acknowledge the financial support obtained from the Swiss National Science Foundation through grant 10.001.670, from grant MICIU/AEI/10.13039/501100011033, and from FEDER, UE. We are grateful to Pol Antras, Jordi Blanes, Riccardo Camboni, Wouter Dessein, Florian Englmaier, Rachel Kranton, Jin Li, Gerard Padro-i-Miquel, Brian Silverman, Pablo Spiller, Paola Valbonesi, Yasutora Watanabe and audiences at HEC Lausanne, IESE, the University of Hong Kong, Sogang University, Ewha Womens University, Waseda University, University of Tokyo, the Paris-Berkeley Organizational Economics Workshop, the Workshop on Relational Contracts, and the Workshop on Global Trade, Political Economy and Public Procurement for helpful comments.

†Tufts University, silke.forbes@tufts.edu

‡IESE Business School, rgils@iese.edu

§HEC Lausanne, University of Lausanne, giorgio.zanarone@unil.ch

1 INTRODUCTION

Ever since Coase (1937) and Williamson (1971), the "make-or-buy" dilemma has been a central question in the social sciences: should firms internally produce their inputs and services, or should they outsource production to independent providers?

In recent decades, research on this question has been dominated by an "asset ownership view" of firm boundaries. According to theoretical models in this tradition, vertical integration transfers productive assets from the seller to the buyer along with the decision rights (Williamson (1985), ch. 6; Grossman and Hart (1986); Hart and Moore (1990)) and payoff claims (Williamson, 1979; Holmstrom and Milgrom, 1991) attached to those assets. Several studies have provided empirical support for the ownership view, showing that vertically integrated buyers use the assets they control to make unscheduled adaptation decisions (Forbes and Lederman, 2009, 2010), and that integration reduces misalignment between buyers and sellers by muting the latter's incentive to maximize the value of upstream assets (Baker and Hubbard, 2003, 2004; Hansman et al., 2020).

An alternative approach to firm boundaries, which we label the "employment view" for lack of a better term, emphasizes that vertical integration changes *who employs the agents* responsible for input production. Building on Simon (1951), theories in this tradition argue that integration transfers exclusive monitoring and contracting rights over productive agents from the seller to the buyer (Williamson, 1975; Masten, 1988; Holmstrom and Milgrom, 1994; Holmstrom, 1999). Thus, while asset-ownership theories of vertical integration compare the behavior of buyer's employees to that of "owner-operators" (drivers owning their trucks, captains owning their boats, distributors owning their stores), employment theories compare the behavior of non-owners who are employed by different firms (the buyer under integration, the seller under

non-integration). In these models, vertical integration may improve employees' performance because it enables the buyer to directly manage them via tools (i.e., monitoring or task assignments) that she could not use with an outsourced supplier's employees. In other words: "The firm uses its right to set internal rules of the game and to design incentives in a manner that internalize some of the contractual externalities that are present in markets" (Holmstrom (1999), p. 76).[1]

The employment view of firm boundaries is appealing because it applies to settings increasingly important in knowledge-based economies, where employees play a greater role than non-human assets in input production. However, empirical analyses of the employment view are rare. Azoulay (2004) and Espinosa (2021) provide evidence (respectively, on pharmaceutical clinical trials and lobbying) where projects are either vertically integrated or outsourced depending on the nature of employees' tasks. Yet, we are not aware of any evidence of how changes in employment relationships across firm boundaries affect firm performance. This paucity of evidence is not surprising: testing the employment view empirically is challenging because both asset ownership and employment relationships are simultaneously modified by vertical integration, making it difficult to isolate the latter's effect on firm performance.

Our paper uses rich data from the US airline industry to tackle this challenge. Both asset ownership and employee performance are important in this industry, and both are affected by the extent to which flights ticketed by a major airline (such

[1]Some analyses of vertical integration emphasize the importance of *both* asset ownership and employment relationships. For instance, transaction cost theory argues that vertical integration unifies asset ownership, and that this reduces the parties' opportunities and motivation to engage in wasteful rent-seeking or "haggling" (Williamson, 1979, 1985) . At the same time, transaction cost scholars also argue that employees are managed differently from contractors and that different bodies of law apply to disputes between employees of the same firm and those between separate firms (Masten, 1988; Williamson, 1991). Similarly, in proposing his "subeconomy" view of the firm, Holmstrom argues that centralized asset ownership complements the buyer's exclusive monitoring and contracting rights as a tool to create a balanced incentive system (Holmstrom and Milgrom, 1994; Holmstrom, 1999). See Gibbons (2005a) and Dessein (2014) for critical reviews of the theoretical literature on boundaries.

as American Airlines or Delta Airlines) are operated by the major or a fully-owned subsidiary (vertical integration) or by an independent regional partner (outsourcing). On the ownership side, the major can decide which aircraft to use on vertically integrated flights whereas the regional partner has such decision right on outsourced flights. Most importantly, the major can freely reallocate the airport resources assigned to a vertically integrated flight (gates and especially the flight's landing slot) to optimize network performance during slot rationing episodes (also known as Ground Delay Programs, or GDPs). This adaptation capability is a key driver of service quality and airline reputation in inclement weather months (Forbes and Lederman, 2009, 2010; Gil et al., 2022).

On the employment side, a flight's timeliness depends on the actions and efforts of multiple employees, from the plane's pilot and crew to the staff performing ground service tasks. For instance, a flight is less likely to be delayed if the pilot and crew members show up on time at the departure gate and/or are willing to serve on flights exceeding their contractual quota during congested days. Additionally, a flight is less likely to be delayed when the crew boards and disembarks passengers efficiently and when ground personnel quickly guide the aircraft into and out of the gate before departure and after arrival. When a flight is vertically integrated, the major employs its pilot and cabin crew and often also the ground staff. In contrast, when the flight is outsourced, the pilot and cabin crew, and sometimes the ground staff, are employed by the operating partner.

Using flight-level data on all US airports in the month of February 2017, we estimate performance differences between vertically integrated and outsourced flights *while controlling for differences in asset ownership* at the major and route level. This allows us to isolate performance differences that are attributable to differences in how employees are managed under the two governance forms and are, therefore, not due

to differences in asset ownership. In our main exercise, we estimate how a flight's arrival delay is impacted by whether it is operated by the major airline or a major-owned subsidiary, or outsourced to an independent regional partner. Thanks to the granularity of our data, we can tightly control for the channels through which the major's asset ownership may lead, directly or indirectly, vertically integrated flights to outperform outsourced ones.

A first asset ownership channel is control over aircraft choice. While a major assigns planes to vertically integrated flights in a way that maximizes its own profits, an independently owned regional partner that also serves other airlines on the same route may not always have an incentive to do so. We hold the effect of aircraft choice constant by including plane size and plane type fixed effects as control variables in our regressions, thereby comparing the performance of vertically integrated and outsourced flights that use similar planes.

Another potential asset ownership channel is the control over landing slots and terminal gates, which may lead to performance differences between vertically integrated and outsourced flights in two ways. First, the major can freely use a vertically integrated flight's slot to reduce another flight's delay during GDPs, whereas it must obtain the regional partner's consent to reallocate an outsourced flight's slot. To control for this channel, we use daily data on landing slot exchanges obtained from the Federal Aviation Administration to interact the effect of vertical integration with dummies for GDP days and separate dummies for whether a particular flight was part of a slot exchange on such days.

Second, the major may choose to primarily use vertically integrated flights on certain routes, such as those characterized by frequent adverse weather episodes, to gain the ability to flexibly reassign landing slots as needed (Forbes and Lederman, 2009). Flights on these high-integration routes are both more likely to be vertically

integrated and to land on time (because they benefit from the major's flexibility in the allocation of landing slots on the route) than the flights on low-integration routes. To control for the major's route integration strategy, we include major-route fixed effects in our regressions, exploiting the fact that even in highly integrated (highly outsourced) routes, a given major typically outsources (directly operates) some flights.

A last asset ownership channel that may lead vertically integrated flights to outperform outsourced ones is priority in the assignment of scarce airport-level resources controlled by the major (primarily, other flights' gates and landing slots). If the major considers vertically integrated flights to be more valuable (for instance, because they carry more passengers on larger aircraft, or because some passengers scrutinize the identity of the operating carrier and believe the major to be more reliable than an outsourced partner), it may give priority to those flights in the assignment of the airport landing slots and gates it controls during GDPs and other emergencies. This, in turn, may lead vertically integrated flights to be on average less delayed than outsourced ones. To mute this channel, we control for various proxies for a flight's value to the major (the plane number of seats and the average fare per major, route and organizational form) in our regressions. The major is the residual claimant on ticket revenues on all flights, including outsourced ones, because it pays a flat fee to its outsourcing partners.

After holding constant a major airline's ability and willingness to allocate flight-relevant assets across vertically integrated and outsourced flights, we find that vertically integrated flights are significantly less delayed than otherwise similar outsourced ones. We estimate that on normal days (i.e., without GDPs and slot exchanges), the average delay of major-operated flights is about six minutes lower than the delay of outsourced flights that are of similar size and value, ticketed by the same major, departing on the same day and at the same time, and flying the same route and

landing at the same airport. The difference in delays across organizational forms becomes even more pronounced on GDP days characterized by adverse weather and slot rationing, on which employee performance matters more at the margin. We find qualitatively similar results when comparing outsourced flights to flights operated by major-owned regional companies (rather than by the major itself), and when we use instrumental variables based on historical weather patterns in the major's network to further control for the endogeneity of vertical integration decisions.

The fact that, even controlling for asset ownership effects, vertically integrated flights are less delayed than outsourced ones indicates that the observed difference must come from improvements in the performance of flight-relevant employees (pilots, plane crew, ground personnel). Our results are therefore consistent with employment models of firm boundaries (Williamson, 1975; Masten, 1988; Holmstrom, 1999): holding asset ownership effects constant, major airlines can elicit greater effort from their own employees, whom they directly monitor and motivate, than from the employees of an outsourced partner.

The rest of this paper is organized as follows. Section 2 discusses our contributions to the literature. Section 3 describes buyer-supplier relationships in the US airline industry. Section 4 presents a simple "employment model" of vertical integration and outsourcing in the airline industry, which generates testable predictions. Sections 5 and 6 present our data, methodology and empirical results. Section 7 concludes.

2 RELATION TO THE LITERATURE

Our paper's primary contribution is to the empirical literature on firm boundaries and the make-or-buy decision. Early research in this literature, reviewed by Lafontaine and Slade (2007), has shown that consistent with classic transaction cost economics

(Coase, 1937; Williamson, 1971, 1979), vertical integration replaces outsourcing when spot markets for inputs are unavailable, and buyers and sellers are locked into bilateral monopoly situations (Monteverde and Teece, 1982; Masten, 1984; Joskow, 1985). More recently, Forbes and Lederman (2009) have shown that consistent with another prediction of transaction cost economics (Williamson, 1991), vertical integration is more frequent when non-contractible ex post adaptation decisions are important. None of these papers attempts to disentangle changes in asset ownership and changes in employment relationships as separate determinants of the vertical integration decision.

More recent studies have tested the importance of the asset ownership mechanism in driving the vertical integration decision. For instance, consistent with Grossman and Hart (1986), Acemoglu et al. (2010) show that the importance of the buyer's specific investments (measure by downstream technology intensity) is positively correlated with vertical integration whereas the importance of the supplier's investments (upstream technology intensity) is negatively correlated with it. Additionally, and consistent with Holmstrom and Milgrom (1991), there is evidence that an agent is less likely to own the assets she uses to work for the principal when her asset-maintenance effort is easier to monitor (Baker and Hubbard, 2004).

A few studies also document how the changes in asset ownership driven by vertical integration affect firm performance. Forbes and Lederman (2010) show that by transferring the ownership of landing slots and airport resources to major airlines, integration enables them to use those resources to reduce flight delays at the airport. In a study of the Peruvian fishing industry, Hansman et al. (2020) show that vertical integration may reduce a supplier's ability to pursue private benefits at the expense of service quality by reducing her discretion to make decisions. Like Forbes and Lederman (2010), we investigate the effect of vertical integration on flights' on-time

performance. However, while they look at how airport-level integration reduces delays of major-operated flights, we hold the major's ownership of airport assets constant and compare the performance of vertically integrated and outsourced flights. This strategy enables us to shed light on a different mechanism through which vertical integration operates – namely, changes in employment relationships.

As discussed in the Introduction, there is very little evidence on the employment mechanism. Azoulay (2004) and Espinosa (2021) study the determinants of integration in settings (respectively, pharmaceutical clinical trials and lobbying) where non-human assets are unlikely to be important. They show that vertical integration of projects is systematically related to the nature of employees' tasks. Our study offers two novel contributions. First, we provide evidence of how changes in employment relationships across firm boundaries affect firm performance. Second, we provide evidence that the employment channel is important even when it coexists with the asset ownership channel, and we develop an empirical strategy to disentangle the two.

3 VERTICAL RELATIONSHIPS IN THE US AIRLINE INDUSTRY

In the U.S. airline industry, major airlines (United Airlines, Delta, American Airlines, Alaska Airlines) operate larger aircraft and outsource flights on smaller aircraft to so-called regional airlines. These regional airlines often serve shorter routes connecting a hub to a local airport. They are sometimes owned by the major but most often are independent companies that partner with the major to operate flights under the latter's banner. Examples of major-owned regionals are Piedmont Airlines or Endeavor Air, while independent partners include Skywest, Trans States Airlines, and Republic Airways.

Relationships between majors and regional partners (both major-owned and independent) are governed by "capacity purchase agreements" (CPAs), under which the regional company operates specific flights under the major's brand, while the major sets flight schedules, sells tickets, and pays the fuel costs for such flights. The major collects all revenues, and the regional receives a flat fee for each operated flight. CPAs provide the regional partner with some incentives for timely performance, such as minimum quotas of timely landings that are necessary to collect the contractual fees, and bonuses linked to on-time performance.

The primary purpose of outsourcing flights to regionals is to reduce the cost of operating those flights, especially because regional airlines can avoid paying the higher, union-negotiated wages and benefits that majors must pay under their labor agreements (Forbes and Lederman, 2009). Further, Forbes and Lederman (2009) and Forbes and Lederman (2010) show that there is a tradeoff between outsourcing to independent vs. major-owned regionals. On the one hand, independent regionals are characterized by greater labor cost discipline. On the other hand, major-owned regionals display better operational performance.

It is important to stress that CPAs are by no means exclusive outsourcing agreements – a given major typically relies on multiple regional partners, even on the same route, and independent regional airlines are often large companies that serve multiple majors. Moreover, while formal CPAs are boilerplate and very similar across major-regional pairs, there is considerable variation, even within a major, in the length and value of outsourcing relationships, and thus in the major's ability to enter relational, self-enforcing agreements with its partner (Gil et al., 2022).

Two distinctions are crucial to understanding the tradeoffs between different governance forms in this paper. The first distinction (normal days vs. GDP days) regards the severity of the contracting frictions that may threaten a flight's on-time

performance. The second distinction (alienable vs. inalienable decisions) regards the different kinds of actions that may matter for flight performance, and how these actions are differently affected by the flight's governance form.

3.1 Normal Days vs. GDP Days

In days of inclement weather, additional time between landings is required to ensure aircraft and passenger safety. At busier airports, this safety requirement prompts the Federal Aviation Administration (FAA) to launch a Ground Delay Program (GDP) that reduces the number of allowed landings per hour. GDPs are frequent events, occurring in almost three quarter of days during winter months, and are concentrated at the more congested airports. GDPs cause both direct and indirect increases in flight delays, relative to days of normal airport operations.

Direct Effect: Because a GDP introduces a binding constraint on the authorized number of landings, it directly forces airlines to delay or cancel several flights. To reduce the resulting passenger inconvenience and reputation damage, major airlines seek to optimally reschedule their flights by submitting slot exchange requests to the FAA, as described in Gil et al. (2022) and Argyres et al. (2025). However, slot exchanges are challenging to implement, and may therefore not always succeed, for two reasons. First, when and where a GDP will hit and which slots will need to be exchanged is hard to predict *ex ante*. Second, once the GDP occurs, agreement over the necessary slot exchanges must be reached under severe time pressure, leaving little or no time for any negotiation between the affected airlines.

Indirect Effect: slot scarcity implies that during a GDP, flights' on-time performance is less resilient to small perturbations and operational hiccups than during normal days. Suppose, for instance, that a flight departs ten minutes late due to a technical problem or slow staffing of the plane. In a non-GDP day, the flight's

landing delay may not be much greater than the departure delay because absent slot rationing, authorities at the destination airport may allow that flight to land ten minutes after the originally scheduled time. Under a GDP, however, if the flight misses its scheduled landing time, slot rationing may prevent it from landing for hours and may even force its cancellation (unless the flight benefits from a slot exchange).

GDPs can, therefore, be both a generator and an accelerator of flight delays.

3.2 Alienable vs. Inalienable Decisions

Airlines can take several actions to reduce flight delays, both on normal days and during GDPs. Some of these actions concern the use of assets (e.g., aircraft, assigned gates, assigned landing slots). How to use these assets is an *alienable decision* in the sense that its control can be transferred by selling the assets (Gibbons, 2005a). For instance, an airline uses planes from its own fleet for the flights it operates, and thus has residual control over which aircraft to use for each flight. Most importantly, a flight's operating airline owns its landing slot and thus has residual control rights over the use of that slot.

In addition to being affected by airlines' alienable decisions, a flight's performance also depends on the efforts exerted by several airline employees (pilots, cabin crew, ground staff) to ensure timely departure and landing. These efforts are *inalienable decisions* because they are delegated to specialized employees regardless of who owns the airline operating a flight. Employees can be monitored and incentivized by airlines but retain the ultimate discretion of how much effort to exert. Employees effort can importantly affect flight performance, especially on GDP days.

3.3 The Role of Firm Boundaries

Whether a flight is vertically integrated or outsourced can affect its performance through both the alienable and inalienable decisions discussed above. The effect of boundaries on alienable decisions is unlikely to come from the relationship-specific investments emphasized by transaction cost and property rights theories (Williamson, 1971, 1979; Grossman and Hart, 1986) because neither physical assets nor human capital are highly specific in the airline industry (Forbes and Lederman, 2009).

Instead, vertical integration directly affects the adoption of those alienable decisions, such as slot exchanges during GDPs, that are difficult to negotiate and agree upon contractually across boundaries. If a flight is vertically integrated, ownership of the slot enables the major to execute a slot exchange involving that flight without frictions – for instance, by assigning the slot to another, more valuable flight ticketed by the major. If instead the flight is outsourced to a regional partner, the major must obtain the partner's consent in order to execute the exchange. If the partner disagrees (due to the personnel and logistical costs of delaying or canceling the flight), the major cannot enforce the slot exchange.

Vertical boundaries can also affect airline employees' motivation to exert effort, although through a different mechanism from the asset ownership transfer one through which they affect alienable decisions. While airline employees own no assets regardless of firm boundaries, changes in boundaries do affect who employs and manages these agents. Pilots and cabin crew are employees of the major in major-operated flights whereas they are employees of the regional partner in outsourced flights. Moreover, at airports where independent regionals have ground service operations, both flight and ground services may be outsourced for certain flights, in which case firm boundaries also determine who employs the ground personnel.

While the major can directly monitor, manage and incentivize its own employees, it can only indirectly affect how an outsourced regional's employees are managed (through the timely performance quota and other incentives contained in the Capacity Purchase Agreement). Thus, incompleteness of the CPA may translate to lower motivation and effort of outsourced regionals' employees, relative to the major's employees, as implied by employment theories of the firm (Masten, 1988; Holmstrom, 1999). Our empirical analysis below aims to assess whether changes in employment relationships affect inalienable decisions, holding the effect of asset ownership on alienable decisions constant.

Notice that while we have emphasized the dichotomous distinction between vertical integration and outsourcing, there are two ways in which a flight can be vertically integrated – namely, via major operation or operation by a major-owned regional company. The major has substantial control over both alienable decisions and employee management at major-owned regional airlines, given that it owns their assets and can appoint and dismiss their top management. For this reason, we expect the effect of vertical integration via a major-owned regional on flight performance to be directionally similar to the effect of vertical integration via major operation. At the same time, the major's control over flights operated by an owned subsidiary is somewhat weaker than the control it exerts over major-operated flights, due to the delegation of managerial decisions to a separate entity. Because of that, in our empirical exercises below we separately analyze how the two types of vertical integration affect flight performance, relative to outsourcing.

4 MODEL

In this section we develop a simple model, adapted from Holmstrom (1999), that illustrates how vertical integration affects the performance of airline employees and motivates our empirical analysis. While Holmstrom (1999) jointly models changes in asset ownership and employment relationships, we solely focus on employment here because our empirical exercises aim to isolate the employment channel while hold the ownership one constant.

Consider a flight ticketed by major airline M. The flight's value to M, $y = fa$, is determined by the effort $a \in \mathfrak{R}^+$ exerted by a representative agent A (e.g., a pilot) at cost $c = \frac{a^2}{2}$. This agent may be employed by M (integration) or by an independent regional partner R (non-integration). All players are risk-neutral.

The agent's efficient (first best) effort maximizes total surplus $y - c = fa - \frac{a^2}{2}$, and is given by $a^{FB} = f$. The question we ask is which contractual arrangement gets M closest to reaching this efficient benchmark. We consider both input-based employment contracts, whereby A's effort is monitored and A receives a stipulated payment if and only if she exerts the agreed upon effort, and output-based contracts, which condition payments on some realized outcome that depends on effort and thus do not require monitoring.

As in Holmstrom (1999), our employment model of firm boundaries has two key ingredients: (a) there are contracting frictions, and (b) different contractual arrangements are available within and between firms. Regarding (a), we follow standard agency theory in assuming that A's individual contribution to M's profit, y, is unverifiable by courts and hence non-contractible (Gibbons, 2005b). A contractible but noisy proxy for A's effort exists and is given by the flight's on-time performance $p = fa + \epsilon$, where ϵ is a zero-mean random variable. However, we assume incentive

contracts based on p are constrained by limited liability: while M has deep pockets, A and R cannot make positive upfront payments (Sappington, 1983).[2] Limited liability is of course only one of the many frictions that may constrain contracts in this industry. It will soon become clear that modeling alternative frictions, such as multitasking and gaming (Holmstrom and Milgrom, 1991), would generate qualitatively similar results.

Regarding ingredient (b), we follow Masten (1988) and Holmstrom (1999) in assuming that contracting possibilities depend on who employs A: while M can monitor A, and hence enter an input contract with her, under integration, only R can enter such contract under non-integration because an employer has exclusive monitoring and contracting rights over its employees (see also Williamson (1991)).[3] Thus, all M can do under non-integration is offer an output contract to R in the hope to influence R's own contract with A, leading to inefficiencies that we analyze below.

4.1 Employee Performance Under Integration

Under vertical integration, M can monitor A and contract with her on effort. We assume M can make a take-it-or-leave-it offer to A, implying that for a given targeted effort a, the optimal input contract requires M to offer a salary equal to the effort cost $c = \frac{a^2}{2}$. A's limited liability implies that M obtains higher profits from an input contract than from an output-based one: M contracts for the effort level that solves

$$max_a \quad \{y - c = fa - \frac{a^2}{2}\},$$

[2]Regional airlines are typically smaller than major airlines and enjoy lower profit margins. Notice that while we assume full liability for simplicity, the model's results would be unchanged if A and R were allowed to make payments up to a low enough ceiling.

[3]The employer's exclusive rights also imply that M cannot offer an output contract to A under non-integration. It will soon become clear, however, that even if feasible, such a contract would not add value.

which results in the first best effort $a^{VI} = f = a^{FB}$.

4.2 Employee Performance Under Non-Integration

Under non-integration, R can monitor A and contract with her on effort, whereas M can only contract with R on the verifiable output p. We focus on linear incentive contracts of the form $w = bp$, where w is the payment that M offers to R and b is a commission rate.[4] Given w, R designs the input contract with A to solve

$$max_a \quad \{E[w] - c = bfa - \frac{a^2}{2}\},$$

which results in the effort schedule $a^*(b) = bf$. Anticipating R's choice, M designs the output contract to solve

$$max_b \quad \{y - E[w] = fa^*(b) - bfa^*(b) = bf^2 - b^2f^2\},$$

which results in the output sharing rule $b^* = 1/2$ that elicits effort $a^{NI} = a^*(b^*) = \frac{f}{2} < a^{VI}$. Moreover, the effort reduction under non-integration increases in f, the effort's marginal productivity: $a^{VI} - a^{NI} = \frac{f}{2}$. The non-integration effort is lower than the integration one because non-integration constrains the set of contractual instruments M can use to elicit effort. Being unable to monitor and contract with A, M must use an output contract to incentivize R to manage A. Given R's limited liability, in order to elicit a given effort a, this output contract must leave R a rent above the cost of effort, which in turn induces M to elicit less effort than would be efficient(Sappington, 1983).[5]

[4]Non-linear contracts may overcome the limited liability problem under certain conditions (Innes, 1990). However, non-linear contracts may introduce other distortions that render them impractical (Gibbons, 2005b). Moreover, linear contracts have been shown to be optimal in certain settings (Holmstrom and Milgrom, 1987).

[5]From the effort schedule $a^*(b) = bf$ we can compute the piece rate necessary to elicit a given effort a, which is $b(a) = \frac{a}{f}$. The compensation M must pay A to elicit effort a is therefore $b(a)E[p] = a^2$,

Proposition. *Employee effort, and the flight's expected on-time performance, are higher under vertical integration than under non-integration:* $a^{VI} > a^{NI}$. *The more productive the effort (the higher* f*), the larger the increase in employee performance under integration.*

The key takeaway from our model is that integration increases effort and the flight's value, relative to non-integration, not because it transfers decision and payoff rights to M (asset ownership is fixed, and the only productive decision in the model, effort choice, inalienably belongs to A), as in ownership theories of firm boundaries (Grossman and Hart, 1986; Hart and Moore, 1990; Williamson, 1971). Instead, integration increases effort because it allows M to use incentive tools (monitoring and input contracts) that would not be available under non-integration (Masten, 1988; Williamson, 1991; Holmstrom, 1999).

In the next section, we bring the predictions of our model to data.

5 DATA AND EMPIRICAL METHODOLOGY

5.1 Data

To conduct our empirical study, we combined four different databases. The first one comes from the Official Airline Guide (OAG hereafter), which provides information on a flight's operating and ticketing airline, flight number, airport, day and scheduled landing hour for all flights landing at a US airport during the winter quarter of 2017.

The second dataset is the On-Time Performance data from the BTS, which provides information on delays for all domestic US flights, except for flights operated by small domestic airlines, mostly major-fully-owned and independent regional airlines. We complement the publicly available information from the BTS with data

which is twice the effort cost.

from FlightAware, which contains delay information from flights excluded in the BTS dataset.

Our fourth and last dataset reports all instances in which a flight is rescheduled to facilitate another flight's landing during airport-based GDPs mandated by the FAA. We obtained this information for all U.S. airports during February 2017 by submitting a Freedom of Information Act request to the FAA.[6]

We measure (the inverse of) a flight's on-time performance, $Delay$, as the difference between its arrival at the gate and the scheduled arrival. We then use the OAG data to measure the governance form of each flight. Specifically, we construct a dummy for whether a flight is directly operated by the major, VI, and a dummy for whether the flight is operated by a regional partner that is fully owned by the major, VIP.

We also construct dummies for whether a flight occurs in a day in which a GDP is enacted at the airport, GDP, and a second dummy, SE, for whether conditional on a GDP being in place, the flight is included in a slot exchange package. Lastly, we construct several control variables to be included in our regressions. Specifically, we measure a flight's number of seats ($Seats$) and the route's average ticket price in the previous year divided by the route distance ($Fare\ per\ Mile$), and we construct dummies for both the aircraft type used to operate the flight. We present summary statistics for all variables in Table 1.

[TABLE 1 HERE]

Table 1 shows summary statistics for our full sample of 537,753 flights (top panel) as well as subsamples for flights occurring on days when there is GDP at the airport

[6]More details on the construction of the FAA database and on slot exchanges can be found in Argyres et al. (2025), which uses a subset of this data (on the exchanges between majors and outsourced partners) to study how adaptation is influenced by a major's decision to use few versus multiple partners.

(mid panel) and flights included in slot exchanges (bottom panel). It is easy to see that the average delay in the full sample is 5.7 minutes, but this increases to 41.8 minutes during GDPs and 74.3 minutes for flights involved in slot exchanges. Note that there the number of seats or fare per mile are quite similar across samples, supporting the exogeneity of GDP occurrences and the wide range of flights included in slot exchanges.

5.2 Empirical Methodology

The goal of our paper is to study how a flight's governance affects its performance. To do so, in our baseline exercises we estimate by OLS linear regression models of the following type:

$$\begin{aligned} Delay_{f,m,r,a,d} \;=\; \alpha + \beta_1 \cdot \mathbf{G}_{f,m,r,a,d} + \gamma_1 \cdot GDP_{a,d} + \gamma_2 \cdot SE_{f,m,r,a,d} + \delta \cdot \mathbf{X}_{f,m,r,a,d} \\ + \lambda_{m,r} + \varepsilon_{f,m,r,a,d} \end{aligned} \tag{1}$$

$Delay_{f,m,r,a,d}$, the dependent variable, measures the delay (in minutes) with which flight f ticketed by major m, serving route r (directional connection between two airports) and landing at airport a on day d. $G_{f,m,r,a,d}$ is a vector of indicators for the flight's governance (more precisely defined below). $GDP_{a,d}$ is an indicator for whether there is a GDP at airport a on day d. $SE_{f,m,r,a,d}$ is an indicator for whether the flight is part of a slot exchange package submitted by the major. $\mathbf{X}_{f,m,r,a,d}$ is a vector of independent variables, including the controls and interactions between independent variables and the governance indicators. The error term $\varepsilon_{f,m,r,a,d}$ is assumed to be normally distributed and iid.

To ensure that our estimates of the effect of governance on performance are not driven by the major's ownership of slots at the airport, and to control for major

airline capabilities and route characteristics (such as weather, airport congestion and passenger traffic) that may affect flight performance, we exploit the granularity of our data and include major-route fixed effects ($\lambda_{m,r}$) in our regressions.

Additionally, we include day-of-month fixed effects and fixed effects for the flight's departure hour at the airport of origin to ensure that our results are not driven by the endogenous selection of flights scheduled on certain days or hours into specific governance forms, and to control for any features of February 2017 days that may affect traffic and airport congestion on and around a flight, such as holidays and weekday-specific patterns. In some regressions we also include plane type fixed effects to control for the possibility that planes with different degrees of efficiency may be systematically used under different governance forms.

6 RESULTS

6.1 Estimating How Vertical Integration Affects Performance After Controlling for Asset Ownership

In our first exercise, we estimate regression equation (1) using VI and VIP – our dummies for whether a flight is operated by the major or by a major-owned regional, respectively – as the governance variables. We present our econometric estimations in Table 2 below.

[TABLE 2 HERE]

Our analysis shows that vertically integrated flights are systematically less delayed than outsourced flights. Columns 2 and 5 show that overall, delays of major-operated flights are about 7 minutes shorter than for outsourced flights. Similarly, delays of flights operated by major-owned regionals are between 1.6 to 3.6 minutes shorter than

those of outsourced flights. This implies that, out of the three governance forms, the on-time performance of major-operated flights is best, outsourced flights perform worst, and the performance of major-owned regionals falls in between.

Columns 3 and 6 show that the performance advantage of vertically integrated flights is similar to the overall average on non-GDP days, in which operations at the landing airport are normal but technical issues or disruptions at the departure airport may still cause landing delays.

Columns 4 and 7 show that on days with a GDP, when routine landing procedures are severely disrupted and employee efforts matter even more at the margin, overall delays increase for all airlines. Our key interest is in the performance difference across governance forms on those days. We find that for flights that are not part of a slot exchange, the difference in delays between major-operated flights and outsourced flights grows by more than 4 minutes (VI*GDP interaction). Controlling for slot exchanges is important because when a flight is included in a slot exchange, its delay tends to either decrease (if it is the slot's recipient) or increase (if it is being delayed or canceled to make space for another flight) by design. Moreover, the major's ability to include a flight in a slot exchange is affected by vertical integration via the major's ownership of the slot. We also find that major-operated flights in a slot exchange have delays that are 17 minutes shorter than those of outsourced flights in slot exchanges (VI*SE interaction). Lastly, we find that flights operated by major-owned regionals have similar performance advantages over outsourced flights (VIP*GDP and VIP*SE interactions).[7]

In Columns 5-7, we also control for aircraft type, i.e. the type of asset that is

[7]Appendix Table A1 shows that our results are robust to restricting attention to the sample of flights operated by major airlines only, namely, American Airlines, Delta Airlines, United Airlines and Alaska Airlines. These are the only US airlines outsourcing flights to independent regional airlines.

deployed. The results are quite robust to including this control. In appendix Table A3, we additionally restrict the sample to smaller planes of less than 143 seats (the 25th percentile of plane size) to ensure that non-linear differences in flight performance are due to aircraft size. This allows us to focus on the set of major-operated planes that is most similar to those used by regionals. Our results are robust to analyzing this part of our sample.

[TABLE 3 HERE]

In Table 3, we control for the number of seats and its interactions with the VI and VIP governance dummies. We use seats as a proxy for a flight's value to the major, which may affect the major's willingness to prioritize that flights in the allocation of scarce resources it owns (primarily, landing slots and gates at the airport). We find that flights with more seats have shorter delays when the flight is outsourced whereas for vertically integrated flights, increasing the number of seats does not significantly affect delays because the interactions VI*Seats and VIP*Seats are of roughly the same magnitude as the effect of Seats, leading the total effect to be statistically insignificant.

When controlling for seats, the effect of vertical integration on performance in non-GDP days tends to be smaller than in Table 2, especially for larger planes. For example, taking the results from Column 2 and using the median plane size for regional flights (65 seats), the difference in arrival delays between major-operated and outsourced flights is major-owned regional and outsourced flights is equal to $-19.8 + 0.285 * 65 = -1.275$ minutes, which is not statistically different from zero in this case. At 50 seats, which is the 25th percentile of the seat distribution of regional flights, the difference would increase to $-19.8 + 0.285 * 50 = -5.55$ minutes.

However, even after controlling for plane size and plane type (Columns 4-6), the

performance effect of vertical integration on GDP days remains virtually unchanged compared the previous table. For these flights, we continue to find large delay reductions when they are vertically integrated (both major-operated and major-owned regional) compared to when they are outsourced. The direct effects of GDP and SE likewise remain very similar to the previous table. Altogehther, these findings show that in days in which airport operations are disrupted, the observed performance improvements under vertical integration are not impacted by flight value as measured by plane size.

Appendix Table A2 shows that our results also hold when we measure flight value by the average fare per mile. In that table, we use the average fare per mile for the carrier on the route in the previous year, to avoid endogeneity concerns. Lastly, appendix Table A4 shows that the results are robust to restricting attention to planes smaller than 143 seats.

Overall, these findings support the employment view of vertical integration illustrated by our model. Holding the major's direct control over decisions affecting flight on-time-performance (namely, choice of plane, and the use of slots and airport resources) constant, and controlling for the major's willingness to allocate resources to a particular flight (through flight size and value controls), vertically integrated flights are less delayed than otherwise identical outsourced flights. Moreover, the effect is stronger in GDP days, when employees' efforts are more crucial at the margin due to the disruption of landing schedules and the scarcity of airport resources.

Because the lower delay of vertically integrated flights does not come from an increase in the major's control rights, it must come from improvements in the incentives of flight-relevant agents (pilots, cabin crew, and ground personnel). Moreover, these agent performance improvements cannot be driven by changes in the agents' asset ownership because pilots, cabin crew and ground personnel are employees, not

owner-operators, regardless whether they work for the major or for an outsourced regional partner. Our results are therefore consistent with the view that integration improves employee performance by giving the "buyer" (in our case, the major) superior monitoring and contracting rights over those employees (Williamson, 1975; Masten, 1988; Holmstrom, 1999).

6.2 Endogeneity of Governance Choices

Majors do not randomly choose whether to operate a flight themselves, assign it to a major-owned regional, or outsource it to an independent regional. Thus, one may worry that despite our rich set of controls and fixed effects, our results may be biased by the endogeneity of governance forms. In discussing endogeneity concerns, we must distinguish between two decisions made by major airlines (Forbes and Lederman, 2009). First, majors choose whether to operate a flight using large aircraft or small regional jets. If the major chooses to use a large aircraft, contracts with the unions require the major to operate that flight directly. Second, when a major has chosen to use a regional jet to operate a flight, it must decide whether to assign that flight to a major-owned regional or to an outsourced regional partner.

The first decision (operating a flight through a large aircraft or through a regional jet) is not driven by the endogenous matching of governance form to transaction costs. This decision tends to be driven by market size or distance between airports. Nevertheless, one may worry that majors schedule major-operated flights at hours and on routes where there is more demand for on-time flights, and schedule regional-operated flights when consumers are more price sensitive and willing to exchange on-time arrivals for lower prices. The regressions presented in Tables 2 and 3 address this potential concern by using major*route fixed effects to control for the "place" endogeneity dimension, and departure airport*departure hour fixed effects to account

for time-of-day variations in the price-sensitivity of demand at a given airport. We have also included day-of-February fixed effects to account for days of the week, weekend, and week-within-a-month cycles in travellers' willingness to pay.

The second decision (assigning a flight operated by a regional jet to a major-owned regional vs. independent regional partner), however, may be driven by the major's optimal matching of governance form to transaction costs, and one might worry that our fixed effects may not fully control for this problem. In this section, we use an instrumental variable approach to further reassure that our results are not biased by the endogenous choice between a vertically integrated partner and an independent regional.

Following the empirical strategy in Forbes and Lederman (2009) and (2010), we use average historical weather - specifically historical precipitation levels - at connecting airports in the major's network as an instrument for the major's decision to vertically integrate a flight. Holding constant the actual amount of precipitation on the day of the flight, historical precipitation levels at airports connecting to the focal flight's landing airport will affect the major's use of vertically integrated flights on those routes. This is because organizational form choice decisions are medium-term decisions that are not adjusted daily in response to current weather conditions. The historical precipitation at connecting airports, in turn, influences the airline's choice of organizational form on the focal route because it is cost-effective to align organizational form choices within the airline's network for connected routes that can share the same aircraft and crew.

We present two sets of instrumental variables (IV) estimations. First, we restrict the sample to planes at or below the 25th percentile of seats (that is, 143 seats or less), and keep observations for all three organizational forms (major's operation, operation by a major-owned regional or operation by an outsourced regional partner).

Because we have only a single instrument - historical precipitation amounts in the network - we construct a single organizational form choice variable that we instrument for. This variable combines major-operated and major-owned regional flights into a single category which we call "vertically integrated". We restrict attention to smaller planes because, as discussed above, the decision to operate large planes (all of which are owned by majors) is not driven by transaction cost considerations. Thus, we exclude the larger major-operated planes from our IV exercise.

In the second set of IV regressions, we restrict the sample further to only include flights operated by regional airlines. These airlines exclusively operate planes called regional jets, which are the smallest category of aircraft. In these regressions, we therefore only compare major-owned regionals to independent regionals and instrument for the integration decision (the VIP variable defined above) using the same historical precipitation instrument as in the first set of IV regressions.

Note that our instruments vary at the major-route level. Therefore, we cannot estimate these regressions with major-route fixed effects as we have done in the previous tables. Instead, we use major-origin airport and major-destination airport fixed effects to control for the major's ownership of airport resources. Similarly, because we do not have sufficient variation in the data to include airport-specific departure hour fixed effects, we replace them with a single set of fixed effects for the departure hour at all airports. We maintain the day-of-the-month fixed effects as before.

[TABLE 4 HERE]

Table 4 shows the IV results for the sample of flights with up to 143 seats which includes major-operated flights. The first two columns show first OLS and then IV results for the effect of vertical integration on arrival delays, controlling for fixed effects for day, departure hour, and carrier-origin and carrier-destination combinations.

Consistent with Table 2, the OLS estimate shows that vertically integrated flights have, on average, delays that are 6.4 minutes shorter than outsourced flights. In the IV regression of Column 2, this performance advantage increases to 12.8 minutes.

In Column 3, we use our IV approach to estimate how these effects are impacted by GDPs. We control for the GDP day and slot exchange dummies as well as their interactions with the vertical integration dummy. After instrumenting for governance choice, we do not find statistically significant differences in arrival delays between integrated and outsourced flights on non-GDP days. However, our findings show that vertical integration still matters on GDP days, when employee effort is more important at the margin. Specifically, we find that the coefficient of the interaction between vertical integration and GDP is about 42 minutes, indicating that integrated flights that are not included in a slot exchange have much shorter delays than similar outsourced flights. This difference appears to be primarily driven by the substantially longer delays of outsourced flights on GDP days (36.6 minutes longer on average, relative to non-GDP days). In contrast, we observe no statistically significant change in arrival delays for vertically integrated flights on GDP days compared to non-GDP days as the point estimate is $36.6 - 42.7 = -6.1$, which is negative but no statistically different from zero.

In the remaining regressions in this table, we control for route distance and the actual amount of precipitation in the network on that day. We add these controls because our fixed effects in these regressions are no longer at the airline-route level, as we explained above. Controlling for current network precipitation ensures that our instrument (historical network precipitation) is not biased by potential correlation between historical and contemporary weather (Forbes and Lederman, 2010). Including these additional controls has no significant impact on our other variables of interest.

Finally, we control for three different measures of the value of the flight: the

number of seats (column 5), the average fare (column 6) and average revenue (column 7) for the airline on this route in the previous year. These controls also have little impact on our other variables of interest. We find that higher flight value as measured by seats or average fare in the prior year has a negative impact on delays. This indicates that airlines prioritize higher-value flights over lower-value ones and likely give additional resources to these flights that allow them to have shorter delays.

[TABLE 5 HERE]

Table 5 presents the IV results for the sample of regional flights (excluding major-operated ones). Here, we find again that, after instrumenting, both the overall effect of a flight being operated by a major-owned regional (VIP coefficient in Column 2) and the effect on non-GDP days (coefficient in Column 3) are not statistically significant. However, vertical integration continues to matter on GPD days. In fact, our results show that outsourced flights have delays that are about 36 minutes longer compared to non-GDP days, whereas the delays of major-owned regional flights are shorter by about 30 minutes. When we control for proxies of the flight's value, we find negative impacts of the number of seats and the revenue in the prior year, again indicating that airlines prioritize higher-value flights. Controlling for the flight's value has a minor impact on the other variables of interest.

It is worth noticing the change of the vertical integration coefficients in our IV regressions (Tables 4 and 5), relative to the corresponding OLS regressions (Table A3, which restricts attention to major-operated flights of less than 143 seats). On the one hand, the stand-alone coefficients for vertical integration switch from negative and statistically significant values in Table A3 to positive and non-significant values in Tables 4-5. On the other hand, the interactions between vertical integration and the GDP (and SE) indicators in Tables 4-5 are larger in magnitude than the OLS

interactions in Table A3 and statistically significant. While the former change in coefficients is the expected correction towards zero from instrumenting, we believe the latter change reflects the existence of heterogeneous effects of our IV estimation. In other words, the larger negative interaction coefficients in Tables 4 and 5, relative to the OLS estimations, are due to the estimation of Local Average Treatment Effects (LATE). Our coefficients are capturing the effect for the specific group of flights that are truly affected by the instrument – that is, those whose vertical integration status changes because of the instrument.

Overall, the instrumental variable results show that vertically integrated flights perform better than outsourced ones on GDP days when airline operations are severely disrupted and employee efforts matter more at the margin. These findings cannot be attributed to changes in decision control and incentives induced by asset ownership, or to the major's decision to vertically integrate into more valuable flights. However, our results are consistent with our hypothesis that by changing the employment status of flight-relevant agents (from employees of an independent regional to employees of a major's subsidiary), integration enables the major to better monitor and incentivize those agents.

7 CONCLUSION

This paper has investigated an understudied rationale for vertical integration – namely, enhancing a buyer's ability to manage and motivate its supplier unit's employees (Holmstrom, 1999). We have presented a simple model that illustrates this "employment view" of firm boundaries in our empirical setting, the U.S. airline industry. The model shows that when interfirm contracts are imperfect, vertical integration enhances the effort of employees who are critical to flights' on-time performance. It

does so by enabling the major airline (the buyer) to directly manage those employees, rather than contracting a regional partner (the supplier) to manage them on her behalf.

Testing the employment view of vertical integration is challenging because of the paucity of employee-level data and because integration changes not only employment relationships but also the ownership of assets critical to flight performance (planes, landing slots, and gates). Our empirical strategy is thus to (a) estimate the effect of vertical integration on performance after controlling for all potential ownership channels, and (b) attribute that residual effect to changes in how employees are motivated and managed within versus across boundaries. Using daily data on flights ticketed by all U.S. major airlines across all airports in February, 2017, we have shown that consistent with the employment view, flights ticketed by a given major are less delayed if operated by the major herself than if operated by an outsourced regional partner, *holding constant the major's ability and willingness to allocate assets the two types of flights*.

While this study focuses on a specific setting, its implications are potentially broad. Vertical integration has been often explained by firms' need to secure the control of critical physical assets and protect relationship-specific investments from holdup (Grossman and Hart, 1986; Williamson, 1979). While these problems may become less important in knowledge-based economies centered on the interplay between human capital and AI, the management and motivation of employees will continue to be a central issue in those settings. By showing that vertical integration has performance effects beyond asset ownership, our results therefore suggest that it should continue to be a key part of firms' strategic governance design toolkit.

REFERENCES

Acemoglu, D., Aghion, P., Griffith, R., and Zilibotti, F. (2010). Vertical integration and technology: Theory and evidence. *Journal of the European Economic Association*, 8:989–1033.

Argyres, N., Gil, R., and Zanarone, G. (2025). Does supply concentration encourage cooperation? evidence from airlines. *Management Science*, 71:5083–5105.

Azoulay, P. (2004). Capturing knowledge within and across firm boundaries: evidence from clinical development. *American Economic Review*, 94:1591–1612.

Baker, G. and Hubbard, T. (2003). Make versus buy in trucking: Asset ownership, job design, and information. *American Economic Review*, 93:551–572.

Baker, G. and Hubbard, T. (2004). Contractibility and asset ownership: On-board computers and governance in u.s. trucking. *Quarterly Journal of Economics*, 119(4):1443–1479.

Coase, R. (1937). The nature of the firm. *Economica*, 4:386–405.

Dessein, W. (2014). Incomplete contracts and firm boundaries: New directions. *Journal of Law, Economics and Organization*, 30:i13–i36.

Espinosa, M. (2021). Labor boundaries and skills: The case of lobbyists. *Management Science*, 67:1586–1607.

Forbes, S. and Lederman, M. (2009). Adaptation and vertical integration in the airline industry. *American Economic Review*, 99(5):1831–1849.

Forbes, S. and Lederman, M. (2010). Does vertical integration affect firm performance? evidence from the airline industry. *RAND Journal of Economics*, 41(4):764–790.

Gibbons, R. (2005a). Four formalizable theories of the firm. *Journal of Economic Behavior and Organization*, 58:200–245.

Gibbons, R. (2005b). Incentives Between Firms (and Within). *Management Science*, 51:2–17.

Gil, R., Kim, M., and Zanarone, G. (2022). Relationships under stress: Relational outsourcing in the u.s. airline industry after the 2008 financial crisis. *Management Science*, 68(2):1256–1277.

Grossman, S. and Hart, O. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94:691–719.

Hansman, C., Hjort, J., Leon-Cillotta, G., and Teachout, M. (2020). Vertical integration, supplier behavior and quality upgrading among exporters. *Journal of Political Economy*, 128(9):3570–3625.

Hart, O. and Moore, J. (1990). Property rights and the nature of the firm. *Journal of Political Economy*, 98:1119–1158.

Holmstrom, B. (1999). The firm as a subeconomy. *Journal of Law, Economics & Organization*, 15:74–102.

Holmstrom, B. and Milgrom, P. (1987). Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica*, 55:303–328.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics & Organization*, 7:24.

Holmstrom, B. and Milgrom, P. (1994). The firm as an incentive system. *American Economic Review*, 84:972–991.

Innes, R. (1990). Limited Liability and Incentive Contracting with Ex Ante Action Choices. *Journal of Economic Theory*, 52:47–65.

Joskow, P. (1985). Vertical integration and long-term contracts: The case of coal-burning electric generating plants. *Journal of Law, Economics and Organization*, 1:33–80.

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

Masten, S. (1984). The organization of production: Evidence from the aerospace industry. *Journal of Law and Economics*, 27:403–417.

Masten, S. (1988). A legal basis for the firm. *Journal of Law, Economics and Organization*, 4:181–198.

Monteverde, K. and Teece, D. (1982). Supplier switching costs and vertical integration in the automobile industry. *The Bell Journal of Economics*, pages 206–213.

Sappington, D. (1983). Limited liability contracts between principal and agent. *Journal of Economic Theory*, 29:1–21.

Simon, H. (1951). A Formal Theory of the Employment Relationship. *Econometrica*, 19:293–305.

Williamson, O. (1971). The vertical integration of production: Market failure considerations. *American Economic Review*, 61:112–123.

Williamson, O. (1975). *Markets and Hierarchies*. New York: Free Press.

Williamson, O. (1979). Transaction cost economics: The governance of contractual relations. *Journal of Law and Economics*, 22:233–261.

Williamson, O. (1985). *The Economic Institutions of Capitalism*. New York: Free Press.

Williamson, O. (1991). Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly*, 36:269–296.

TABLES

Table 1. Summary Statistics of Main Variables in Full Sample, GDP and SE subsamples

	No. Obs	Mean	Std. Dev.	Min	Q1	Q2	Q3	Max
Arrival Delay	537,753	5.666	44.780	-60	-16	-6	10	768
VI (major-operated)	537,753	0.631	0.483	0	0	1	1	1
VIP (operated by major-owned regional)	537,753	0.104	0.306	0	0	0	0	1
Major	537,753	0.730	0.444	0	0	1	1	1
GDP at airport	537,753	0.142	0.350	0	0	0	0	1
SE (flight in slot exchange)	537,753	0.073	0.260	0	0	0	0	1
Seats	537,753	121.184	50.189	30	76	143	160	366
Fare per mile	520,079	0.479	0.200	0.067	0.331	0.447	0.592	4.662
If GDP = 1								
Arrival Delay	76,610	41.801	72.921	-60	-7	17	68	767
VI (major-operated)	76,610	0.673	0.469	0	0	1	1	1
VIP (operated by major-owned regional)	76,610	0.054	0.226	0	0	1	1	1
Major	76,610	0.764	0.424	0	1	1	1	1
SE (flight in slot exchange)	76,610	0.512	0.500	0	0	1	1	1
Seats	76,610	127.791	51.333	30	76	143	160	366
Fare per mile	74,565	0.466	0.228	0.087	0.302	0.405	0.569	1.928
If SE = 1								
Arrival Delay	39,242	74.254	79.240	-59	17	52	112	761
VI (major-operated)	39,242	0.650	0.477	0	0	1	1	1
VIP (operated by major-owned regional)	39,242	0.044	0.206	0	0	0	0	1
Major	39,242	0.810	0.392	0	1	1	1	1
Seats	39,242	125.904	51.750	50	76	143	160	366
Fare per mile	38,462	0.481	0.227	0.102	0.317	0.426	0.576	1.928

This table provides summary statistics for all variables used in our analysis in the full working sample, sample of flights arriving to airports under GDP and a sample used in slot exchanges.

Table 2. Effect of Vertical Integration on Flight Performance (Full Sample)

	1	2	3	4	5	6	7
Dependent Variable:	**Arrival Delay (Minutes)**						
VI (major-operated)	-6.380***	-6.929***	-7.452***	-5.533***	-7.015***	-7.034***	-4.072***
	(0.734)	(0.641)	(0.565)	(0.546)	(1.665)	(1.363)	(1.170)
VIP (operated by major-owned regional)	-4.471***	-1.588**	-1.361**	-0.206	-3.567***	-3.392***	-2.357***
	(1.239)	(0.655)	(0.576)	(0.529)	(0.678)	(0.598)	(0.542)
GDP at airport			9.724***	13.077***		9.713***	13.395***
			(1.607)	(2.567)		(1.607)	(2.565)
SE (flight in slot exchange)			57.155***	69.403***		57.128***	69.536***
			(3.630)	(5.346)		(3.631)	(5.344)
VI*GDP				-4.050**			-4.469**
				(1.868)			(1.874)
VI*SE				-17.472***			-17.704***
				(4.002)			(4.002)
VIP*GDP				-5.803**			-6.283**
				(2.654)			(2.636)
VIP*SE				-14.242*			-14.492*
				(8.468)			(8.455)
Constant	10.158***						
	(1.320)						
FE Day of Month	NO	YES	YES	YES	YES	YES	YES
FE Major-Route	NO	YES	YES	YES	YES	YES	YES
FE departure hour-Origin Airport	NO	YES	YES	YES	YES	YES	YES
FE General Aircraft Type	NO	NO	NO	NO	YES	YES	YES
Observations	537,645	537,645	537,645	537,645	537,645	537,645	537,645
R-squared	0.004	0.206	0.282	0.284	0.207	0.283	0.285

Variable definitions
vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3. Controlling for Flight Value, Proxied by Number of Seats (Full Sample)

	1	2	3	4	5	6
Dependent Variable:	**Arrival Delay (Minutes)**					
VI (major-operated)	-31.871***	-33.162***	-32.523***	-18.577***	-20.444***	-17.587***
	(2.437)	(2.220)	(2.173)	(2.961)	(2.679)	(2.618)
VIP (operated by major-owned regional)	-18.876***	-19.812***	-20.766***	-16.291***	-17.218***	-17.151***
	(2.107)	(2.087)	(2.088)	(2.073)	(2.051)	(2.066)
GDP at airport		9.719***	13.260***		9.709***	13.379***
		(1.606)	(2.567)		(1.607)	(2.567)
SE (flight in slot exchange)		57.179***	69.554***		57.142***	69.587***
		(3.631)	(5.352)		(3.632)	(5.348)
VI*GDP			-4.280**			-4.441**
			(1.873)			(1.875)
VI*SE			-17.658***			-17.761***
			(4.004)			(4.004)
VIP*GDP			-6.163**			-6.375**
			(2.655)			(2.645)
VIP*SE			-14.375*			-14.483*
			(8.470)			(8.462)
Seats	-0.314***	-0.332***	-0.351***	-0.181***	-0.204***	-0.205***
	(0.027)	(0.024)	(0.024)	(0.031)	(0.029)	(0.029)
VI*Seats	0.338***	0.353***	0.372***	0.160***	0.184***	0.186***
	(0.028)	(0.025)	(0.025)	(0.032)	(0.030)	(0.029)
VIP*Seats	0.267***	0.285***	0.318***	0.208***	0.227***	0.242***
	(0.033)	(0.032)	(0.033)	(0.033)	(0.032)	(0.033)
FE Day of Month	YES	YES	YES	YES	YES	YES
FE Major-Route	YES	YES	YES	YES	YES	YES
FE departure hour-Origin Airport	YES	YES	YES	YES	YES	YES
FE General Aircraft Name	NO	NO	NO	YES	YES	YES
Observations	537,645	537,645	537,645	537,645	537,645	524,824
R-squared	0.282	0.284	0.207	0.283	0.285	0.306

Variable definitions

vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Effect of Vertical Integration on Flight Performance, Instrumental Variables Regressions (Planes <=143 Seats)

	1	2	3	4	5	6	7
	OLS	IV	IV	IV	IV	IV	IV
Dependent Variable	**Arrival Delay (Minutes)**						
Vert Int (VI + VIP)	-6.393***	-12.840**	-6.244	9.268	17.797*	7.858	9.650
	(0.267)	(6.380)	(6.430)	(5.954)	(10.680)	(6.310)	(6.209)
GDP			36.612***	36.791***	36.828***	36.829***	36.889***
			(4.014)	(3.995)	(3.991)	(3.998)	(3.995)
SE (flight in slot exchange)			100.767***	100.877***	100.494***	101.128***	101.030***
			(6.116)	(6.083)	(6.085)	(6.101)	(6.077)
Vert Int * GDP			-42.707***	-43.118***	-43.189***	-43.171***	-43.271***
			(6.918)	(6.883)	(6.874)	(6.889)	(6.884)
Vert Int * SE			-82.265***	-82.312***	-81.293***	-82.789***	-82.636***
			(12.089)	(12.001)	(12.018)	(12.039)	(11.988)
Route Distance				-0.000	0.000	0.001	-0.000
				(0.001)	(0.000)	(0.001)	(0.001)
Same Day Precip, Connect Airports				86.960***	84.666***	87.716***	87.492***
				(12.871)	(13.196)	(12.937)	(12.805)
Seats					-0.215*		
					(0.120)		
Ln(Fare)						-1.679***	
						(0.596)	
Ln(Revenue)							-0.523
							(0.371)
FE Day of Month	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carrier-Dep Airport FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carrier-Arr Airport FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	319,848	319,848	319,848	319,848	319,848	319,848	319,848
Mean of Dep Var	6.893	6.893	6.893	6.893	6.893	6.893	6.893

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

Variable definitions

vert int = dummy for whether flight is either operated by major or by major-owned regional (that is, vert int = VI + VIP)

(the omitted reference category is "flight operated by outsourced regional")

GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.

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

Table 5. Effect of Vertical Integration on Flight Performance, Instr Var Regressions (Regional Flights Only)

	1	2	3	4	5	6	7
	OLS	IV	IV	IV	IV	IV	IV
Dependent Variable	**Arrival Delay (Minutes)**						
VIP	-3.837***	-4.649	7.500	5.952	7.934	6.032	7.530
	(0.371)	(11.891)	(11.393)	(11.505)	(11.229)	(11.622)	(11.364)
GDP			35.648***	35.846***	35.853***	35.784***	35.955***
			(3.201)	(3.207)	(3.203)	(3.213)	(3.214)
SE (flight in slot exchange)			60.758***	60.680***	60.745***	60.703***	60.749***
			(2.989)	(2.992)	(2.989)	(2.994)	(2.992)
VIP * GDP			-65.152***	-65.657***	-65.616***	-65.503***	-65.933***
			(7.824)	(7.842)	(7.824)	(7.855)	(7.856)
VIP * SE			3.583	3.755	3.681	3.748	3.595
			(8.470)	(8.475)	(8.461)	(8.474)	(8.473)
Route Distance				-0.006***	-0.004***	-0.004**	-0.005***
				(0.001)	(0.001)	(0.002)	(0.001)
Same Day Precip, Connect Airports				80.842	71.227	80.445	72.946
				(49.530)	(49.163)	(49.694)	(49.433)
Seats					-0.303***		
					(0.046)		
Ln(Fare)						-1.111	
						(1.743)	
Ln(Revenue)							-0.949***
							(0.182)
Observations	132,004	132,004	132,004	132,004	132,004	132,004	132,004
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dep Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carrier-Dep Airport FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Carrier-Arr Airport FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep Var	8.220	8.220	8.220	8.220	8.220	8.220	8.220

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

Variable definitions

vip = dummy for whether flight operated by major-owned regional

(major-operated flights are excluded from the sample so the omitted reference category is "flight operated by outsourced regional")

GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.

Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A1. Effect of Vertical Integration on Flight Performance (Flights Ticketed by Major Airlines)

	1	2	3	4	5	6	7
Dependent Variable:	**Arrival Delay (Minutes)**						
VI (major-operated)	-7.387***	-7.030***	-7.621***	-5.822***	-33.773***	-35.357***	-34.814***
	(0.747)	(0.611)	(0.548)	(0.527)	(2.565)	(2.351)	(2.310)
VIP (operated by major-owned regional)	-4.472***	-1.623**	-1.355**	-0.252	-18.541***	-19.885***	-20.770***
	(1.239)	(0.649)	(0.568)	(0.526)	(2.106)	(2.087)	(2.080)
GDP at airport			9.768***	13.094***		9.755***	13.273***
			(1.848)	(2.569)		(1.847)	(2.569)
SE (flight in slot exchange)			59.947***	69.674***		59.986***	69.840***
			(3.922)	(5.305)		(3.923)	(5.311)
VI*GDP				-4.560***			-4.827***
				(1.763)			(1.770)
VI*SE				-15.357***			-15.565***
				(4.019)			(4.017)
VIP*GDP				-5.680**			-6.050**
				(2.642)			(2.641)
VIP*SE				-13.894*			-14.009*
				(8.369)			(8.369)
Seats					-0.311***	-0.334***	-0.353***
					(0.027)	(0.024)	(0.024)
VI*Seats					0.350***	0.368***	0.387***
					(0.029)	(0.026)	(0.026)
VIP*Seats					0.261***	0.286***	0.318***
					(0.033)	(0.032)	(0.032)
Constant	9.432***						
	(1.307)						
FE Day of Month	NO	YES	YES	YES	YES	YES	YES
FE Major-Route	NO	YES	YES	YES	YES	YES	YES
FE departure hour-Origin Airport	NO	YES	YES	YES	YES	YES	YES
Observations	392,665	392,665	392,665	392,665	392,665	392,665	392,665
R-squared	0.005	0.213	0.292	0.293	0.213	0.292	0.294

Note: This table restricts attention to flights that are ticketed by major airlines (American Airlines, Delta Airlines, United Airlines and Alaska Airlines)

Variable definitions
vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2. Controlling for Flight Value, Proxied by Fare per Mile

	1	2	3	4	5
Dependent Variable:	**Arrival Delay (Minutes)**				
VI (major-operated)	-6.355***	-2.785	-7.796***	-6.809***	-5.137**
	(0.719)	(2.100)	(2.477)	(2.098)	(2.107)
VIP (operated by major-owned regional)	-4.312***	-0.411	6.962***	6.420***	7.946***
	(1.234)	(2.891)	(1.987)	(1.816)	(1.861)
Fare per mile		5.672			
		(4.166)			
VI*Fare per mile		-6.168*	1.824	-0.890	-0.441
		(3.290)	(4.040)	(3.481)	(3.523)
VIP*Fare per mile		-6.523*	-13.561***	-12.350***	-12.934***
		(3.694)	(2.803)	(2.729)	(2.727)
GDP at airport				9.763***	13.191***
				(1.619)	(2.567)
SE (flight in slot exchange)				57.037***	69.146***
				(3.625)	(5.310)
VI*GDP					-4.183**
					(1.863)
VI*SE					-17.200***
					(3.993)
VIP*GDP					-5.843**
					(2.654)
VIP*SE					-14.066*
					(8.393)
Constant	10.037***	6.667**			
	(1.318)	(2.834)			
FE Day of Month	NO	NO	YES	YES	YES
FE Major-Route	NO	NO	YES	YES	YES
FE departure hour-Origin Airport	NO	NO	YES	YES	YES
Observations	520,079	520,079	520,009	520,009	520,009
R-squared	0.004	0.004	0.206	0.283	0.285

Variable definitions
vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3. Effect of Vertical Integration on Flight Performance (Planes ≤ 143 Seats)

	1	2	3	4	5	6	7
Dependent Variable:	**Arrival Delay (Minutes)**						
VI (major-operated)	-6.970***	-8.399***	-8.858***	-6.942***	-10.914***	-10.131***	-6.675***
	(0.829)	(0.615)	(0.565)	(0.540)	(2.316)	(1.882)	(1.640)
VIP (operated by major-owned regional)	-4.472***	-2.224***	-1.940***	-0.798	-4.337***	-4.089***	-3.060***
	(1.239)	(0.676)	(0.588)	(0.544)	(0.700)	(0.611)	(0.554)
GDP at airport			11.084***	13.712***		11.068***	13.910***
			(1.853)	(2.614)		(1.852)	(2.615)
SE (flight in slot exchange)			60.322***	69.178***		60.341***	69.281***
			(4.414)	(5.334)		(4.416)	(5.335)
VI*GDP				-3.413*			-3.789*
				(2.068)			(2.082)
VI*SE				-19.756***			-19.939***
				(3.981)			(3.987)
VIP*GDP				-6.615**			-6.979***
				(2.633)			(2.617)
VIP*SE				-14.256*			-14.480*
				(8.407)			(8.397)
Constant	10.158***						
	(1.320)						
FE Day of Month	NO	YES	YES	YES	YES	YES	YES
FE Major-Route	NO	YES	YES	YES	YES	YES	YES
FE departure hour-Origin Airport	NO	YES	YES	YES	YES	YES	YES
FE General Aircraft Type	NO	NO	NO	NO	YES	YES	YES
Observations	335,142	335,142	335,142	335,142	335,142	335,142	335,142
R-squared	0.005	0.217	0.292	0.295	0.218	0.293	0.296

Variable definitions
vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A4. Controlling for Flight Value, Proxied by Number of Seats (Planes ≤143 Seats)

	1	2	3	4	5	6
Dependent Variable:	**Arrival Delay (Minutes)**					
VI (major-operated)	-41.831***	-41.046***	-36.635***	-7.119	2.358	6.895
	(5.837)	(5.141)	(4.663)	(19.065)	(17.586)	(17.276)
VIP (operated by major-owned regional)	-19.646***	-20.314***	-20.993***	-16.676***	-17.378***	-17.346***
	(2.127)	(2.103)	(2.107)	(2.136)	(2.102)	(2.112)
GDP at airport		11.071***	13.829***		11.060***	13.914***
		(1.852)	(2.615)		(1.851)	(2.616)
SE (flight in slot exchange)		60.357***	69.298***		60.366***	69.322***
		(4.417)	(5.341)		(4.417)	(5.338)
VI*GDP			-3.617*			-3.779*
			(2.075)			(2.084)
VI*SE			-19.957***			-19.992***
			(3.991)			(3.988)
VIP*GDP			-6.914***			-7.098***
			(2.637)			(2.627)
VIP*SE			-14.368*			-14.442*
			(8.411)			(8.404)
Seats	-0.336***	-0.348***	-0.364***	-0.201***	-0.220***	-0.222***
	(0.026)	(0.024)	(0.024)	(0.030)	(0.029)	(0.029)
VI*Seats	0.419***	0.414***	0.401***	0.011	-0.072	-0.082
	(0.052)	(0.046)	(0.043)	(0.190)	(0.177)	(0.173)
VIP*Seats	0.271***	0.285***	0.314***	0.203***	0.219***	0.235***
	(0.033)	(0.032)	(0.033)	(0.034)	(0.033)	(0.033)
FE Day of Month	YES	YES	YES	YES	YES	YES
FE Major-Route	YES	YES	YES	YES	YES	YES
FE departure hour-Origin Airport	YES	YES	YES	YES	YES	YES
FE General Aircraft Name	NO	NO	NO	YES	YES	YES
Observations	335,142	335,142	335,142	335,142	335,142	335,142
R-squared	0.218	0.293	0.296	0.218	0.294	0.296

Variable definitions
vi = dummy for whether flight operated by major; vip = dummy for whether flight operated by major-owned regional
(vi & vip are mutually exclusive dummies, the omitted reference category is "flight operated by outsourced regional")
GDP = dummy for whether there is GDP at airport*day; SE = dummy for whether flight is in slot exchange package during a GDP.
Robust standard errors clustered at the arrival airport/day of the month level in parentheses. *** p<0.01, ** p<0.05, * p<0.1