DISSERTATION

ECONOMIC ESSAYS ON WILDLIFE-AIRCRAFT CONFLICT IN THE UNITED STATES

Submitted by

Jordan Navin

Department of Economics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2019

Doctoral Committee:

Advisor: Stephan Weiler

Aaron Anderson
David Mushinski
Anita Pena
Stephan Kroll

ABSTRACT

ECONOMIC ESSAYS ON WILDLIFE-AIRCRAFT CONFLICT IN THE UNITED STATES

Wildlife-aircraft conflict poses a substantial economic and safety threat in the United States (US). Dolbeer, Wright, Weller, Anderson, and Begier (2014) estimates direct costs related to wildlife strikes burdened the US economy by approximately $157 million annually between 1990 and 2014. In 1995, the Federal Aviation Administration (FAA) collaborated on a project with the United States Department of Agriculture's (USDA) Wildlife Services to investigate the magnitude and nature of the wildlife strike problem, ultimately resulting in the creation of the National Wildlife Strike Database (NWSD). However, reporting strikes (and associated information, such as repair costs) to the NWSD is not mandatory, and information used to calculate economic damage estimates from wildlife strikes in the US relies on voluntarily reported cost data.

This dissertation focuses on the direct costs of wildlife strikes in the US and the associated disclosure behaviors of large domestic American airlines. Chapter 1 investigates the relationship between the likelihood of voluntary repair cost disclosure after a wildlife-strike event by such airlines and market competitiveness and idiosyncratic firm profits. Results show changes in competitiveness and profitability impact the voluntary disclosure of wildlife-strike repair costs by major US airlines to the NWSD. Chapter 2 similarly examines airline voluntary disclosure accuracy, employing emerging methods from economics and accounting literature that test the accuracy of self-reported data based on a statistical property exhibited by large datasets, known as Benford's Law (de Marchi & Hamilton, 2006; Dumas & Devine, 2000; Nigrini, 1996; Zahran, Iverson, Weiler, & Underwood, 2014). Analogous to Chapter 1, findings indicate the accuracy of repair costs American air carriers report to the NWSD is linked to market competition and profits. Chapter 3 relates to developing a method for interpolating missing repair costs in the NWSD using machine learning techniques. Results show that a neural network outperforms both linear regression and random forest models when predicting out-of-sample data, and furthermore, interpolating missing costs in the NWSD with a neural network delivers an average annual estimate of the direct costs of wildlife strikes in the US that is approximately $75 million, significantly less than prior estimates. Specifically, the neural network approach yields estimates $19 and $82 million lower, respectively, than when using mean cost assignment and Dolbeer et al. (2014)'s reported estimate derived using a variation of the same method.

ACKNOWLEDGEMENTS

I would like to thank my chair, Dr. Stephan Weiler for encouraging and believing in my earliest ideas related to the dissertation, as well as my own skills as a researcher. I would also like to thank all of my colleagues at the National Wildlife Research Center, specifically my committee member, Dr. Aaron Anderson. I would not have been able to complete the project without his daily guidance and assistance. Thank you to all of my other committee members.

DEDICATION

This dissertation is dedicated to my grandfather, Dr. Leo J. Navin. Thank you for being such a wonderful role model throughout my childhood and adult life. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The first recorded wildlife collision with any aircraft occurred September 7, 1905, when Orville Wright's plane collided with a bird while flying over a cornfield in Dayton, Ohio (DeVault, Blackwell, & Belant, 2013). In addition to birds, mammals and other wildlife can pose hazards to aircraft. The first fatality from such an event was reported on April 3, 1912, when Calbraith Rodgers, the first person to fly across the continental United States, was killed in a plane crash with a seagull along the coast of Southern California (DeVault et al., 2013). Despite these early recorded events, strikes with birds and other wildlife were of minimal importance during the early years of flight; damage to only three civil aircraft and two human fatalities were recorded worldwide between 1912 and 1959 (DeVault et al., 2013). However, following the rise of turbine-powered aircraft in the 1960s, the magnitude and frequency of damage from wildlife strikes rose significantly. Between 1960 and 2010, wildlife strikes accounted for the destruction of at least 160 civilian aircraft worldwide. In October 1960, a crash in Boston Harbor, Massachusetts, following the ingestion of over 200 European starlings into three of the plane's four engines resulted in the deaths of sixty-two people. More recently, on January 15, 2009, US Airways flight 1549, carrying 155 passengers, was forced to make an emergency landing in the Hudson River in New York City after Canada geese flew into both engines following takeoff from LaGuardia Airport (DeVault et al., 2013).

There are several reasons for the increase in wildlife-related strikes with aircraft over the past 50 years. First, early piston-driven planes were relatively noisy and slow, so birds could typically avoid such aircraft. Modern aircraft are faster, relatively more quiet, and their engine fan blades are more susceptible to strike damage than traditional propeller blades (DeVault et al., 2013). Second, most older propeller-driven aircraft contained three or four engines, while modern planes typically contain two, thus elevating the risk of multiple-engine damage from wildlife strikes (DeVault et al., 2013; Wenning, Begier,

& Dolbeer, 2004). Third, air travel has generally become more commonplace worldwide (Begier & Dolbeer, 2009; DeVault et al., 2013; Dolbeer & Wright, 2008; Dolbeer et al., 2014; Wenning et al., 2004). DeVault et al. (2013) note that data from the Federal Aviation Administration (FAA) indicate that commercial air traffic in the US has increased from approximately 14 million movements in 1975 to 25 million in 2010. Fourth, the increase in air travel has been complemented by the unprecedented success of wildlife and environmental conservation since the 1960s. Successful programs in both settings have contributed to impressive population increases among species particularly harmful to aircraft, including white-tailed deer, American alligators, Canada geese, double-crescent cormorants, sandhill cranes, American white pelicans, seagulls, falcons, hawks, eagles, and wild turkeys. Furthermore, many of these species have populated suburban and urban settings and are thriving in response to habitat protection in such areas (DeVault et al., 2013).

Driven by the simultaneous increase in both commercial aviation and wildlife species considered hazardous to such activities, aircraft collisions with wildlife pose substantial economic and safety problems (Begier & Dolbeer, 2009; Wenning et al., 2004). Dolbeer et al. (2014) estimates an average of $157 million annually in direct repair costs[1] due to wildlife strikes between 1990 and 2014. Results from Anderson et al. (2015) indicate the average direct cost of a damaging bird strike in the US in 1990-2013 was $225,739, with a maximum cost slightly greater than $40 million. In 1995, the FAA collaborated on a project with the United States Department of Agriculture Wildlife Services (USDA/WS) to obtain additional information related to the magnitude and nature of the wildlife strike issue, ultimately resulting in the creation of the National Wildlife Strike Database. Reporting strikes to the NWSD is not mandatory but is strongly encouraged by the FAA. Data used to calculate economic damage estimates, such as Dolbeer et al. (2014)'s, rely on

[1]Costs resulting from structural damage to the aircraft, which excludes indirect costs, such as losses from delayed flights.

the voluntary disclosure of repair cost information resulting from wildlife strike damage by aircraft operators. For the case of large commercial airlines in the United States, a single airline typically has an administrative representative coordinate with on-staff wildlife managers to disclose a given repair cost to the database. The lack of information disclosure concerning these direct costs biases previous estimates of strike-related damage (Anderson et al., 2015; DeVault et al., 2013; Dolbeer, 2015) as economic wildlife-aircraft damage estimates, such as Dolbeer et al. (2014)'s, have relied on naive empirical methods to interpolate missing information, specifically, variations of mean cost assignment. As of July 2018, approximately 30% of strikes in the NWSD that indicated structural damage to the aircraft *ex-post* of a strike event reported a related repair cost.

This dissertation focuses on the *direct* costs of wildlife strikes in the US and associated disclosure behaviors of specifically American domestic air carriers. The broad motivation for this work is to guide wildlife management policy decisions regarding both strategies to raise the level of knowledge related to the economic burden of wildlife strikes (i.e., the cost disclosure policy), as well as the efficient allocation of finite management resources. In Chapter 1, the relationships between voluntary repair cost disclosure by large American airlines and market competitiveness, as well as idiosyncratic firm profits after a wildlife-strike event, are investigated. One of this chapter's contributions is the combination of two previously disparate datasets, matching damaging strike observations from the NWSD with market- and firm-level characteristics obtained from the Massachusetts Institute of Technology's (MIT) Airline Data Project. The results show the probability of large American airlines disclosing direct repair costs after a wildlife strike event are linked to market competition and profitability. Moreover, I find evidence of an interaction effect between competition and profits that is consistent with the existing empirical literature. These findings directly inform policies managing the economic burden of wildlife strikes, most importantly, the current voluntary disclosure policy. The results also provide insight into the broad theory of the economic firm by examining disclosure behavior outside of capital markets.

In Chapter 2, how the accuracy of voluntarily disclosed direct repair costs by the same group of American airlines varies with respect to market competition and profitability is examined. The statistical property of Benford's Law is used to show differences in accuracy between costs disclosed under the relatively more competitive market setting prior to 2008 and under the less competitive market structure since 2008. Differences in the accuracy of costs disclosed by airlines with an at or above average net operating income in the year of the reported strike are also shown. These findings reinforce those of Chapter 1, namely that firm-level voluntary disclosure behavior is linked to market competition and profitability. This work provides additional insight into firms' voluntary disclosure and informs relevant wildlife management policymakers in the same vein as the first chapter.

Chapter 3 investigates the model performance of random forest and neural network machine learning methods versus the traditional econometric tool of regression when interpolating missing costs in the NWSD. First, it is shown that a neural network is the best of these three models at predicting missing wildlife strike repair costs, offering close to a 6% improvement as measured by root mean squared error when predicting out-of-sample data. This information is beneficial to future researchers who encounter a similar missing values problem and are faced with selecting a model to handle them. Given the realities of such missing values and their effect on decisions regarding wildlife-aircraft conflict at institutions like the FAA and USDA, potential biases in previous estimates should be addressed to ensure the efficiency of future related policies. Predicting missing repair costs with a neural network is also shown to deliver an average annual estimate of the direct costs of wildlife strikes in the US that is approximately $75 million – $19 million lower than when using mean cost assignment and $82 million dollars lower than Dolbeer et al. (2014)'s estimate, which uses a variation of the same method. Highlighting the dangers of using naive empirical methods, this information is relevant to policymakers, who could adjust funding to use resources more efficiently. At the airport level, this research provides a potential tool for program efficiency evaluation by allowing airports to estimate damage based on what's observed in reality instead of using an airport or national average.

Chapter 1

Revealing repairs: Market competition, profitability, and wildlife strike cost revelation in the US domestic airline industry – evidence from the National Wildlife Strike Database

1.1 Motivation

Since the FAA last revisited its wildlife strike disclosure policy in 2009, the domestic airline industry has undergone significant changes in market concentration that have been linked to a decline in industry competitiveness (Peterman, 2014; Shen, 2017). The combined market share of the top four domestic US airlines has risen from 65% in 2010 to 84% in 2015 (Forbes, 2016). This is largely the result of several consolidations within the industry, beginning with the merger of Delta Air Lines and Northwest Airlines in 2008. Subsequent mergers include United and Continental Airlines in 2010, Southwest Airlines and AirTran Airways in 2011, and American Airlines and US Airways in 2013.

A major cost of wildlife strike-related repair cost disclosure by commercial airlines is competitive in nature, such that disclosing repair costs to the public NWSD may provide useful operations information to competitors (Hodson, 2017). One main purpose of this chapter is to examine how the aforementioned shift in market structure has impacted US commercial airlines' voluntary repair cost disclosure behaviors in the event of a damaging wildlife strike.

The motivation for this research is two-fold. First, from a wildlife management perspective, such an examination will provide useful insight to those in charge of future changes to relevant disclosure policies. Second, the context of the changing competitive nature of the domestic airline industry and relevant behaviors of its participants provides a new empirical lens to study existing theories of voluntary disclosure. This chapter also investigates the related relationship between firm profitability and disclosure decisions,as well as the interaction between competition and profits. Prior empirical studies have found conflicting results (Botosan & Stanford, 2005; Dedman & Lennox, 2009; Harris, 1998; Verrecchia & Weber, 2006), but their focus has been restricted to the disclosure of earnings measures in relation to firms' ability to attract investor capital (Botosan & Stanford, 2005; Gelb, Henry, & Holtzman, 2008; Harris, 1998; Verrecchia & Weber, 2006). This work broadens the relevant literature by using a dataset of damaging wildlife strikes from the NWSD appended with firm- and market-specific characteristics to contribute new empirical evidence and, further, by examining disclosure decisions that are presumably unrelated to capital markets – similar to Dedman and Lennox (2009).

The issues examined in this chapter can be viewed more generally as an investigation of how variation in market competitiveness and profitability affect information-related behavioral responses at the firm level. Repair costs reported to the NWSD by one airline can potentially reveal a discrete piece of information to competing airlines regarding a comparative advantage in materials, labor, and other factors. An example offered during the previously described interview was the potential revelation of preferential pricing contracts with certain manufacturers of replacement airplane components (e.g., GE and Rolls Royce) for a given airline. Studying such behavioral responses provides novel and useful insight not only into the economic actions of major American commercial airlines but also, more broadly, economic firm theory.

The remainder of this chapter proceeds as follows. Section 1.2 offers a brief review of, connects this work to the extant related literature and formulates *a-priori* hypotheses for each key relationship of interest. Section 1.3 describes the dataset and employed variables. Section 1.4 details the empirical modeling process, while Section 1.5 presents results and a robustness check. Section 1.6 includes a brief discussion and offers a framework of an agenda for future analyses.

1.2 Background, analytical niches, and core hypotheses

1.2.1 Disclosure and market competition

Most of the theoretical literature surrounding *all* types of voluntary disclosure behavior posits that due to relatively higher costs of disclosure, firms operating in more competitive industries disclose less than those operating in less competitive industries (Armantier & Richard, 2003; Arya & Mittendorf, 2007; Board, 2009; Clinch & Verrecchia, 1997; Darrough, 1993; Verrecchia, 1983).[2] However, some empirical articles find a positive relationship between industry competitiveness and firms' voluntary disclosure behavior, hypothesizing that managers seek to protect the abnormally high profits that arise from operating in low-competition industries (Botosan & Stanford, 2005; Harris, 1998). While the Cournot model of oligopoly offers theoretical support for this hypothesis insofar as high profits are correlated with high levels of industry concentration, it says nothing explicitly about managerial incentives to withhold information in the face of high profits, only that a smaller number of participants result in higher profits for firms, which weakens their arguments. Other empirical work is also critical of such findings (Dedman & Lennox, 2009; Verrecchia & Weber, 2006). Dedman and Lennox (2009) find a negative relationship between industry competitiveness and disclosure decisions regarding both earnings and sales information for a sample of private companies from the United Kingdom. For the case of public firms and the disclosure of performance measures in the US, a similar relationship was found by Verrecchia and Weber (2006).

It should be noted that much of the foregoing literature and this chapter choose to consider competition facing airlines specifically to be Cournot or quantity competition in-

[2]Darrough and Stoughton (1990) note that this prediction may be sensitive to the type of competition, such as threat of entry versus current competition.

stead of Bertrand (price) or monopolistic competition. Under Cournot competition, firms simultaneously commit to a certain quantity of a homogeneous product. Market prices are then determined by the total production of all participating firms. In this framework, firm revenue not only depends on idiosyncratic quantity decisions but also on the decisions of competitors, due to market prices being determined by the aggregate supply of all participants.

Airlines typically pre-commit to a specific number of seats via a flight plan dictating when and where flights will be offered before tickets are sold (Lohatepanot & Barnhart, 2004). Desgranges and Gauthier (2016) and Brander and Zhang (1990) have both used this framework to theoretically and empirically model the US domestic airline industry. Desgranges and Gauthier (2016) models the 2008 Delta-Northwest merger using a Cournot oligopoly model, while Brander and Zhang (1990) find empirical consistency between the domestic airline industry and the Cournot model. Moreover, as discussed by Tirole (1988), monopolistic competition assumes that any price change by one firm has only a marginal effect on the demand of any other firm, which is intuitively inconsistent when discussing the airline industry. Furthermore, 31% of all American airline travelers flew only once in 2015, while 20% of Americans flew twice (Airlines for America, 2016). Moreover, of all the individuals who flew, almost half flew for 'personal leisure' rather than business purposes. These statistics lead one to believe that the portion of travelers who would exhibit any kind of behavior, such as brand loyalty, is sufficiently small that it would not impact the assertion about Cournot made in the chapter. Restated, most American airline passengers only fly once per year. Monopolistic competition assumes if one airline changes their price, these individuals will continue buying from airlines that charge a higher price.

Although the existing empirical results are somewhat conflicting, the leaning of most previous research generally supports the conclusion that companies seek to hide information from competitors by limiting voluntary disclosure. Drawing on this primary conclusion, this chapter's hypothesis regarding the competition-disclosure relationship of interest is as follows:

- *Hypothesis 1.1:* The relationship between market competition and the voluntary disclosure of wildlife-strike-related repair cost information for the US domestic airline industry is *negative*.

1.2.2 Disclosure and profitability

Firms may use the voluntary disclosure of performance measures to overcome adverse selection. Entities with more favorable prospects disclose more information, increasing demand for their debt and equity leverage, thus reducing their capital costs (Dye, 1985; Gelb et al., 2008; Verrecchia, 1983). Gelb et al. (2008) specifically examine the relationship between profits and earnings disclosure by commercial airlines before and after industry deregulation during the Carter administration. Prior to the policy treatment, the authors find no significant relationship between earnings and disclosure, but find a "direct, positive relationship" between airlines earnings and their choice of voluntary disclosure post-deregulation. They explain this result by arguing that prior to deregulation, more profitable airlines avoided the voluntary release of information, fearing political costs in the form of adverse regulatory outcomes. Following deregulation and a significant reduction in the political costs of disclosure, the most profitable airlines offered the highest levels of disclosure, consistent with the hypotheses of Verrecchia (1983) and Dye (1985).

Others argue the relationship between profitability and public firms' disclosure decisions is theoretically ambiguous due to two competing forces. The first is that profitable firms are incentivized to signal their success to potential investors, while the second competing factor is that profitable firms are incentivized to withhold information regarding abnormal profitability to avoid imitation by rivals (Dedman & Lennox, 2009). Based on this ambiguity, Dedman and Lennox (2009) adjust their scope beyond the realm of performance measures and capital market effects, finding a significant negative relationship between profitability and the disclosure of information related to firms' cost of sales, as well as sales themselves. In the context of this result, they hypothesize that firms can better hide their sources of success if they choose to withhold such information, explicitly stating that "[a] company with abnormally low costs would not wish to reveal to rivals such a high level of efficiency is possible ... Similarly, a company with abnormally high sales would prefer to conceal this from rivals in order to prevent it from copying its ... strategies" (Dedman & Lennox, 2009, p.210). From this perspective, the cost of voluntary disclosure is higher for more profitable firms such that more profitable entities have 'more to lose' than their less profitable competitors when revealing information.

Theoretical models regarding the disclosure-profitability relationship are sparse. Modeling commercial airlines' revelation of strategic information using a simple theoretical framework of cost-minimization supports the hypothesis of Dedman and Lennox (2009), insofar as more profitable firms face a relatively higher cost of disclosure. This cost-minimization exercise is formally developed in Appendix A. Congruent with Hypothesis 1.1 and suggestive of a parallel relationship, the link between profitability and voluntary disclosure should be negative (Silberberg, 1978).[3] Considering this chapter's theoretical cost-minimization structure and the previous argument that disclosure is costlier for more profitable firms, the following additional core hypothesis is proposed:

- *Hypothesis 1.2:* The relationship between airline profitability and the voluntary disclosure of wildlife-strike-related repair cost information for the US domestic airline industry is *negative*.

[3]Using this framework versus more traditional profit-maximization is beneficial for two reasons. First, it negates the assumption of airlines behaving as price-takers in the output market. Second, commercial aircraft operators are typically constrained to a particular level of production that is conditional on available resources, making cost-minimization a core target metric and, in competitive markets, the theoretical complement to profit-maximization.

1.2.3 The disclosure and competition-profit interaction

The interaction between competition and profit on firm voluntary disclosure decisions is also investigated in this chapter. Dedman and Lennox (2009) propose an *a-priori* expectation the effect will be positive, arguing that "highly profitable companies have no need to protect their positions by withholding information if they face no competitive threat. Therefore, the association between profitability and [disclosure] may be stronger if the company faces greater competition" (Dedman & Lennox, 2009, p.223). Alternatively stated, at lower levels of industry competitiveness, profit should be expected to matter relatively less to firm disclosure behavior. Empirical tests for this competition-profit interaction effect on disclosure yield a positive, statistically insignificant result. The authors note that their findings are limited by the narrower survey nature of their approach and widespread collinearity among explanatory variables, as well as potential endogeneity.

To the author's knowledge, their article contains the only explicit hypothesis offered to explain the relevant interrelationship between profitability and competition. Moreover, the disclosure behavior in Dedman and Lennox (2009) most closely resembles the behavior studied in this work.[4] In this line of thought, the following hypothesis is offered to relate the interaction effect of competition and profit on airline voluntary disclosure:

- *Hypothesis 1.3:* The competition-profit interaction effect on voluntary disclosure of wildlife-strike-related repair cost information for the US domestic airline industry is *positive*.

1.2.4 Additional evidence: The demand-side costs of disclosure

Although the primary costs of voluntary repair cost disclosure for domestic airlines are rooted in the broad context of 'supply,' considering potential disclosure burdens from the 'demand' side provides reinforcing additional intuition for Hypotheses 1.1, 1.2, and,

[4]The similarity between their work and ours is that non-performance-related information should be significant beyond the scope of capital markets.

thus, Hypothesis 1.3. Consumers accessing the NWSD may use revealed costs as a raw measurement of the severity of strikes across commercial airlines and adjust their behavior based on individual preferences for perceived risk. Moreover, cost information could be published by popular press outlets. For example, Figure 1.1 plots the Google Trends web-search interest index for the topic 'bird strike' from January 2004 to September 2017. Each dotted vertical line represents a separate wildlife strike incident published in popular media outlets. Examples include the collision of a turkey vulture with US Space Shuttle Discovery in July 2005, as well as the more serious emergency landing of US Airways Flight 1459 on the Hudson River in January 2009. Delta Airways Flight 1063's 2012 emergency landing at New York City's JFK International Airport following a bird strike during takeoff underscored these risks. Prior disclosure of repair cost information could reduce passenger revenue through either of the aforementioned demand-driven mechanisms, particularly given the internet era's broad and deep flows of media exposure and, thus, increased risk awareness during these densely media-driven cycles.

Through this demand-side lens, it can be inferred that under relatively less competitive product market conditions, the relevant costs of disclosing would be smaller in magnitude compared to industry settings with higher levels of competition. Hence, if passengers are constrained in their choice set for domestic flight offerings, economic intuition implies that any demand-associated costs of repair information disclosure will be lower given the lack of possible substitutes. Furthermore, if at least one source of relatively more profitable airlines' comparative advantage is founded in this demand context, such firms would again have 'more to lose' when voluntarily disclosing repair costs. This demand-side intuition reinforces this chapter's core hypotheses, making the supply-side-focused empirical analysis a relatively conservative estimate of the underlying incentives for voluntary disclosure.

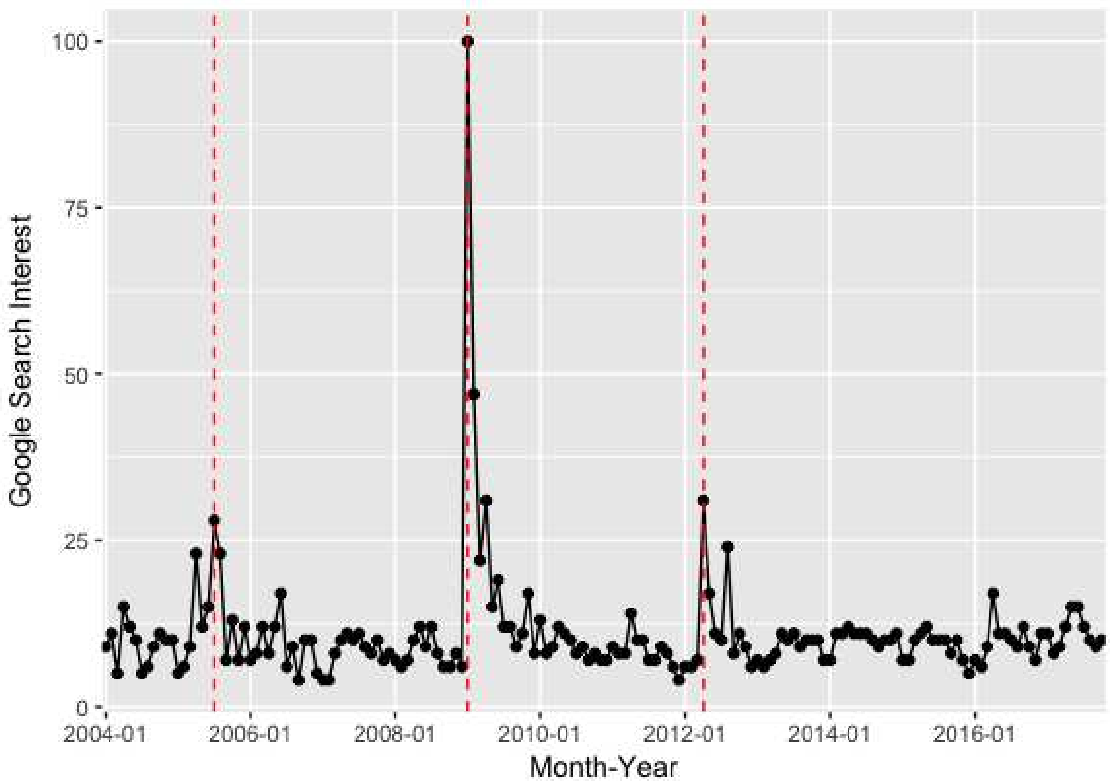


Figure 1.1: Google Trends web search interest for 'bird strike'; 2004-2017

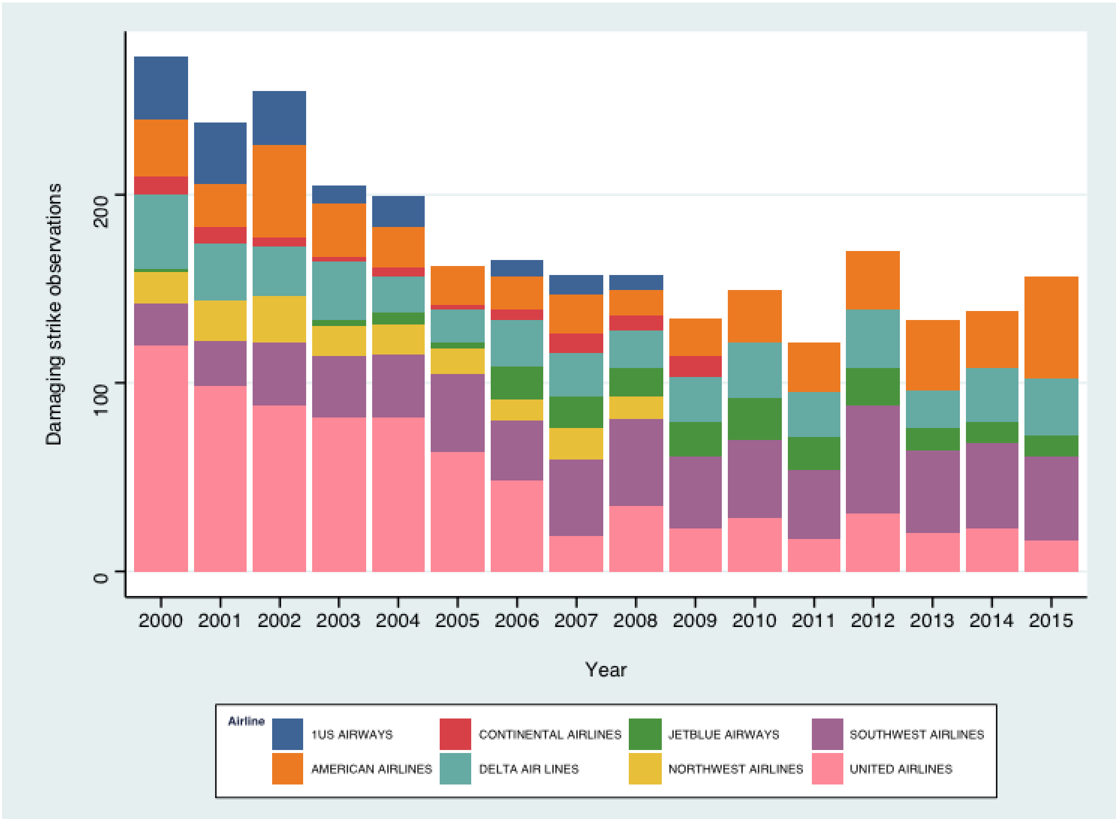
1.3 Data

The NWSD currently contains records of all wildlife strikes in the US involving airplanes, reported on a voluntary basis since 1990 (Dolbeer & Wright, 2009; *Federal Aviation Administration National Wildlife Strike Database*, 2016). The sample used in this chapter consists of 1,476 damaging[5] wildlife strikes from 8 different commercial airline operators spanning the time period 2001-2015, appended with firm-level data from MIT's Airline Data Project[6] (MIT-ADP).[7] Each observation is indexed according to the strike incident, airline, and time. All market- and airline-level variables were taken from or calculated using MIT-ADP data. Figure 1.2 presents an annualized distribution of sample observa-

[5]Structural damage to aircraft was indicated in the initial strike report.

[6]http://web.mit.edu/airlinedata/www/AboutUs.html

[7]All observations consist of strikes that occurred at US airports, restricting the scope to domestic airline operations.



Source: NWSD/MIT–ADP

Figure 1.2: Sample airline composition by year

tions across airlines. Coinciding with increasing market concentration, the number of airlines present in the sample decreases over time. Airlines that merged with another airline do not appear in the sample after the year the merger was finalized. For example, because Continental Airlines merged with United Airlines in 2010, damaging strike observations for Continental do not appear in any subsequent years (i.e., 2010-2015). Across the observed time period, a total of 97,893 wildlife-strike incidents were reported to the NWSD by all aircraft operators. Of these strikes, 38,553 ($\approx$ 39%) were reported by the group of airlines composing the sample. Of the 97,893 total strikes reported, 8,601 ($\approx$ 9%) indicated structural damage to the aircraft. This chapter's sample of 1,476 damaging strikes accounts for approximately 4% of all reported strike incidents by sample airlines. Dolbeer (2015) estimates damaging strike reports in the NWSD account for approximately

78% and 91% of actual occurrences for the time periods 2004-2008 and 2009-2013, respectively.

1.3.1 Primary variables

Table 1.1 presents summary statistics for the three primary variables of interest in this analysis. A binary dependent variable taking a value equal to one if a direct repair cost is present for the damaging strike observation is used to capture airline disclosure behavior. Consistent with the related findings of Anderson et al. (2015), approximately 81.5% of the 1,476 damaging strikes in the sample failed to disclose a related damage cost. The average value of all reported costs in the sample is $318,652. This average was obtained using the 'inflation-adjusted' cost measure provided by the NWSD; however, no reference year is given.

Table 1.1: Select sample summary statistics

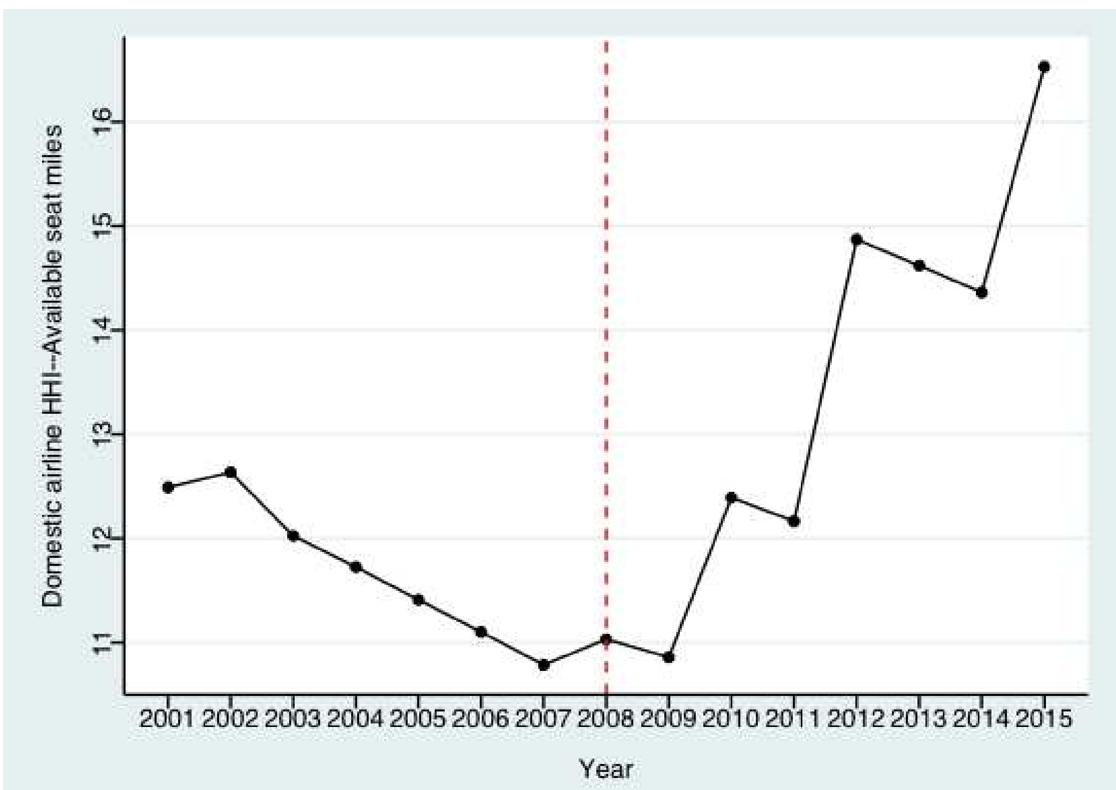
Statistic	Mean	St. Dev.	Min	Max
Cost reported (1=Yes, 0=No)	0.185	0.388	0	1
Herfindahl-Hirschman Index	12.483	1.586	10.786	16.527
Net operating income (USD billions)	0.151	2.245	−8.314	7.802

The Herfindahl-Hirschman Index (HHI) is chosen as the central variable to measure the impact of market competition on disclosure behavior. A statistical measure of concentration, the HHI has been used as a tool by both the US Department of Justice and the Federal Reserve to analyze the competitive effects of significant horizontal mergers (Rhoades, 1993). The HHI's usefulness as a measure of market competitiveness is theoretically grounded in the Cournot model of oligopoly, where firms engage in quantity competition. As the number of firms in a given market increases, the predicted equilibrium outcome trends toward that of a perfectly competitive market. Conversely, as the number of firms falls, the Cournot outcome begins to resemble that of a monopolist, meaning an exit directly implies less competition in the market through lower consumer welfare and higher firm profits. The Cournot model essentially predicts competition to be inversely related to industry concentration.

The HHI is defined mathematically by Tirole (1988) as follows:

$$HHI = \sum_{i=1}^{n} \alpha_i^2$$

Where α_i represents the market share of firm i operating in an industry with n competitors, such that $\alpha_i = q_i/Q$ and $\sum_{i=1}^{n} \alpha_i = 1$. Larger HHI values imply greater industry concentration. Several related studies have used the HHI as a measurement for market competitiveness, including Harris (1998) and Verrecchia and Weber (2006). For meaningful interpretation of empirical results, this analysis multiplies calculated HHI by 100.



Source: MIT Airline Data Project–Domestic Available Seat Miles

Figure 1.3: US domestic airline market share; 2001-2015

Other researchers, such as Dedman and Lennox (2009), have voiced concerns about using industry concentration as a proxy for competitiveness, arguing high concentration does not always translate to low levels of competition as predicted by the Cournot model. Moreover, the concentration-competition relationship can vary across industries, as well as across firms operating in the same product market. While the proposition that high concentration doesn't always lower industry competition may very well be true, in the case of the US domestic airline industry, the concentration-competition relationship predicted by Cournot appears to hold. Examining the 2013 merger between US Airways and American Airlines, Peterman (2014) writes that "the merger will lead to an overall decrease in [industry] competition, which will have anticompetitive effects on the price and quality of air travel" (Peterman, 2014, p.805). Similarly, investigating the impacts of the United-Continental merger in 2010, Shen (2017) finds the price for routes previously competitive between the two airlines significantly rose following the acquisition. Appealing to Cournot and the aforementioned literature, through its use of the HHI, this chapter ultimately assumes that high concentration captures lower market competition for the US domestic airline industry and vice versa. Specifically, the HHI variable used in this analysis is calculated by examining the airline market share of Available Seat Miles (ASMs). ASMs are defined by MIT-ADP as a "common industry measurement of airline output that refers to one aircraft seat flown one mile, whether occupied or not. An aircraft with 100 passenger seats, flown a distance of 100 miles, generates 10,000 available seat miles." Figure 1.3 plots the calculated HHI measure across the observed time period for the sample. An obvious trend shift exists following the Delta-Northwest merger in 2008[8], with the HHI of the US domestic airline industry steadily rising.

Firm profitability is measured using annual net operating income, the difference between operating revenues and expenses.[9] The airline advocacy group Airlines for Amer-

[8]Indicated by the dotted red line.

[9]USD billions

ica defines operating revenues as "[r]evenues from the performance of air transportation and related incidental services, including (1) transportation revenues from the carriage of all classes of traffic in scheduled and nonscheduled services, and (2) nontransportation revenues consisting of federal subsidies (where applicable) and services related to air transportation" and operating expenses as "[e]xpenses incurred in the performance of air transportation, based on overall operating revenues and expenses ... [excluding] non-operating income and expenses, nonrecurring items, or income taxes" (Airlines for America, 2018). The International Air Transport Association notes labor and fuel costs account for $\approx$ 50% on average of expenses for airlines operating in North America in 2001-2008 (IATA, 2010). From a purely economic perspective, the net operating income metric used in this chapter can be viewed as the difference between total revenue and total costs directly related to air transportation. The average of this measure for the sample is approximately $151 million and carries a rather large standard deviation of close to $2.5 billion.

The mean HHI for the sample is approximately 12.5, which the US Department of Justice (DOJ) classifies as "an unconcentrated market." [10] However, since the Delta-Northwest merger in 2008, the industry's HHI rose from 11.1 to 16.5 in 2015, shifting the DOJ's industry classification to 'moderately concentrated.'" The data shows a significant shift in the *relative* market concentration over the sample time period, such that it was sufficient to change its classification status by the DOJ's broad guidelines. Under this DOJ category of 'moderately concentrated,' any horizontal merger that shifts the market of interest's HHI by one unit or more "potentially raise[s] significant competitive concerns and often warrant[s] scrutiny."

[10] https://www.justice.gov/atr/horizontal-merger-guidelines-08192010#5c

1.3.2 Secondary variables

A given strike event in the NWSD contains information on over 90 variables associated with the incident (Anderson et al., 2015). Similar to Anderson et al. (2015), this analysis uses a subset of these variables to control for differences in strike severity and regularity that presumably impact disclosure behavior. The chosen variables include several binary indicators for the aircraft component damaged during a given strike event such that each component is associated with its own indicator for whether it was damaged. These components include the radome, windshield, nose, engine, wing/rotor, fuselage, landing gear, tail, lights, and a category listed in the database as 'other.' Also included is a categorical variable capturing the impact of the damaging strike on the involved aircraft's flight schedule with possible outcomes, including aborted take-off, engine shut down, precautionary landing, 'other,' unknown, and no effect. The phase of the flight at which the strike occurred is controlled for with categorizations including the approach, climb, descent, landing roll, take-off run, taxi, or unknown. The size of the animal struck in the strike incident is controlled for using a categorical variable with possible outcomes of large, medium, and small. The geographical location of the strike event is captured by including a variable indicating the FAA region in which the strike occurred. Possible regions include Alaska (AAL), Central (ACE), Eastern (AEA), Great Lakes (AGL), New England (ANE), Northwest Mountian (ANM), Southern (ASO), and Southwest (ASW). The time of day of the strike event is additionally controlled for by coding the timing as either dawn, day, dusk, night, or 'unknown.' Airport-level fixed effects are included as strike-specific controls to account for variation in reporting behaviors potentially related to the presence of specific competitors at a given airport where a strike event occurs. Detailed summary statistics for all strike-level controls are presented in Table B.1 of Appendix B

The average stage length and daily departures are used as supplementary airline-level controls. An average stage length is the mean time between takeoff and landing and is defined by MIT-ADP as "[t]he average distance flown, measure[d] in statute miles, per aircraft departure. The measure is calculated by dividing total aircraft miles flown by the number of total aircraft departures performed." Daily departures measure the average number of departures of a single aircraft for each respective airline per 24-hour period. Both the average stage length and daily departures are included to control for differences in volume and traffic across airlines, both of which plausibly influence damage cost reporting behavior, as well as profitability, following a wildlife strike. Summary statistics for these variables are located under 'Mkt., oper., and fin. characteristics' in Appendix B Table B.1.

1.4 Empirical modeling

Airline disclosure of repair information is considered costly, because it conveys potentially useful knowledge to competitors. Moreover, in the presence of such competitive costs theory from Section 1.2 indicates profits to be possibly dependent on the level of disclosure chosen by a firm. To account for this potential simultaneity issue, the following equation is empirically estimated for strike event i occurring at time t for airline j using two-stage least squares (2SLS):[11]

$$ReportCost_{ijt} = \alpha_1 + \Phi HHI_{it} + \Omega NOI_{ijt} + \mathbf{X}_i^T \boldsymbol{\beta}_1 + \mathbf{F}_{ijt}^T \boldsymbol{\delta}_1 + \boldsymbol{\Lambda}_{1j} + \boldsymbol{\eta}_{1t} + \epsilon_{1ijt} \qquad (1.1)$$

$ReportCost_{ijt}$ is the binary dependent variable capturing disclosure, taking a value equal to one if a repair cost is reported by airline j following damaging strike i. HHI_{it} is the calculated ASM Herfindahl-Hirschman Index, while NOI_{ijt} is commercial airline j's net operating income metric. Both annual measures are indexed to the year of damaging strike i. $\mathbf{X}_i^T$ is a transposed vector of characteristics specific to strike i, and $\mathbf{F}_{ijt}^T$ is composed of operating characteristics associated with commercial airline j in strike year t that presumably impact both income and reporting behavior: average stage length and daily departures. $\boldsymbol{\Lambda}_{1j}$ is a vector of airline fixed effects and is included to capture time-invariant

[11] Results from a GMM distance test reject exogeneity of NOI_{ijt}

unobservable differences in disclosure behavior across firms. Motivated by an empirical test of lag structure that is presented in Appendix section C, all estimations of (1.1) are performed including (as well as withholding) a one year lag of NOI_{ijt}[12] as an exogenous variable in the system. $\boldsymbol{\eta}_{1t}$ is a vector of year and month fixed effects, and ϵ_{1ijt} is an idiosyncratic error term.

The instrument utilized for 2SLS estimation is airline j's *load factor* in the year (t) of damaging strike event i. A commonly used measure of capacity utilization in the airline industry, a load factor is defined by MIT-ADP as "... the proportion of airline output that is actually consumed." Mathematically, for airline j at year t (and indexed to a single observation, strike i), the load factor is expressed as follows:

$$LoadFactor_{ijt} = \frac{Revenue\,Passenger\,Miles_{jt}}{Available\,Seat\,Miles_{jt}}$$

MIT-ADP defines Revenue Passenger Miles (RPMs) as "[t]he basic measure of airline passenger traffic. It reflects how many of an airline's seats were actually sold." For a given airline, a load factor of zero would imply that every flight offered in that year was completely empty, and conversely, a load factor equal to unity indicates every flight chartered was completely full. Golbe (1986) similarly uses the load factor to capture exogenous variation in airline profitability. Appendix Table B.1 includes summary statistics for the instrument, expressed as a percentage.

This identification strategy assumes that a non-zero correlation between $LoadFactor_{ijt}$ and NOI_{ijt} exists and, further, that $LoadFactor_{ijt}$ is uncorrelated with ϵ_{1ijt}. The latter assumption may be alternatively stated that this model proposes $LoadFactor_{ijt}$ impacts disclosure behavior only through its effect on NOI_{ijt}. Intuitively, (1.1) proposes that an airline's choice of cost disclosure in the event of a damaging wildlife strike is dependent

[12] NOI_{ijt-1}

on industry competitiveness, idiosyncratic profit, characteristics of the damaging strike, several airline-specific operating measures, and time.

Thus, the following first-stage equation is used to obtain predicted values for NOI_{ijt},[13], which are subsequently used to estimate (1.1) and obtain consistent estimates of Ω:

$$NOI_{ijt} = \alpha_2 + \gamma LoadFactor_{ijt} + \phi HHI_{it} + \mathbf{X}_i^T \boldsymbol{\beta}_2 + \mathbf{F}_{ijt}^T \boldsymbol{\delta}_2 + \boldsymbol{\Lambda}_{2j} + \boldsymbol{\eta}_{2t} + \epsilon_{2ijt} \tag{1.2}$$

where, again, $\mathbf{X}_i^T$ is a transposed vector of characteristics specific to strike i, and $\mathbf{F}_{ijt}^T$ is composed of operating characteristics associated with commercial airline j in strike year t. $\boldsymbol{\Lambda}_{2j}$ and $\boldsymbol{\eta}_{2t}$ are airline and month/year fixed effects, respectively. ϵ_{2ijt} is an idiosyncratic error term. (1.2) is essentially a regression of the endogenous regressor, NOI_{ijt}, on all exogenous variables from (1.1) and the identifying instrument $LoadFactor_{ijt}$.

Due to the large number of parameters estimated in the model and the relatively small sample size, (1.1) is estimated using a linear probability model instead of other empirical models for binary dependent variables, such as probit or logit. It should also be noted that the potential collinearity problem that exists in the presence of many parameters and small sample size is acknowledged. However, all variables included in the empirical model are theoretically significant. Omitting them from the model, even though the sample size is small, would be analytically detrimental.

1.5 Estimation results

Table 1.2 presents first-stage results from the estimation of (1.2). Heteroskedasticity-robust standard errors were used in all estimations presented in this chapter. The coefficient $\hat{\gamma}$ associated with $LoadFactor_{ijt}$ is statistically significant at the one percent level, including and withholding the one year lag of net operating income. The negative direction

[13] Denoted $N\hat{O}I_{ijt}$ throughout the remainder of the chapter.

of $\hat{\gamma}$ s counterintuitive; however, such first-stage results provide no direct economic interpretation of any coefficient. Both specifications have a R^2 value above 0.7. Regarding the strength of the employed instrument, tests of overidentification are not possible due to (1.1) being exactly identified. However, the first stage F-statistics, when withholding and including the one year lag of operating income, are 30.70 and 25.13, respectively, more than satisfying Staiger and Stock (1997)'s rule of thumb that this statistic should be greater than 10. Furthermore, the instrument $N\hat{O}I_{ijt}$ is robust to 10% maximal IV size for Stock and Yogo (2005)'s weak instrument test.

Table 1.2: First stage estimation results

	(1)	(2)
	NOI_{ijt}	NOI_{ijt}
$LoadFactor_{ijt}$	-0.182***	-0.163***
	(0.0328)	(0.0326)
HHI_{it}	2.416***	2.267***
	(0.141)	(0.146)
NOI_{ijt-1}	No	Yes
Strike-specific controls	Yes	Yes
Month/year FE	Yes	Yes
Airline FE/operations controls	Yes	Yes
Airport FE	Yes	Yes
N	1476	1476
R^2	0.736	0.745
F-statistic–$LoadFactor_{ijt}$	30.70	25.13

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3 presents average partial effects from 2SLS[14] estimation of (1.1). The significance levels reported for HHI_{it} and $N\hat{O}I_{ijt}$ represent one-tailed hypothesis tests for each associated coefficient. The null hypotheses are $H_0 : \hat{\Phi} \leq 0$ and $H_0 : \hat{\Omega} \geq 0$ for all specifications in the chapter, mirroring Hypotheses 1.1 and 1.2. Confirming Hypotheses 1.1

[14]Using predicted values for NOI_{ijt} obtained from the first-stage equation (1.2).

and 1.2, the coefficient estimate associated with HHI_{it} is positive[15], while the estimated coefficient for the net operating income instrument $\hat{NOI}_{ijt}$ is negative. Both coefficients are statistically significant at the one percent level[16] and hold a relatively consistent magnitude across specifications.

Table 1.3: IV average partial effects: competition, profit

	(1)	(2)
	$ReportCost_{ijt}$	$ReportCost_{ijt}$
HHI_{it}	0.123***	0.125***
	(0.0779)	(0.0833)
$\hat{NOI}_{ijt}$	-0.121***	-0.127***
	(0.0422)	(0.0477)
NOI_{ijt-1}		0.0106
		(0.0135)
Strike-specific controls	Yes	Yes
Month/year FE	Yes	Yes
Airline FE/operations controls	Yes	Yes
Airport FE	Yes	Yes
N	1476	1476

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The DOJ's guidelines on horizontal mergers in relation to industry state that a unit increase in HHI is likely to come under scrutiny from regulators and, thus, is unlikely to actually be observed in reality. A more useful interpretation of the coefficient $\hat{\Phi}$ is in the context of a 0.5 unit increase as a single instance of a unit increase in HHI is unlikely to ever be observed. Thus, the estimated coefficients in Table 1.3 are scaled to represent a 0.5 unit change. Through this lens, a 0.5 unit increase in industry HHI in the year of damaging strike i lowers the likelihood that airline j discloses a repair cost

[15]Recall that HHI is inversely related to market concentration; this chapter assumes an *increase* in concentration represents a decrease in competition, assuming the airlines are engaging in quantity competition.

[16]P-values are equal to 0.0008 (HHI) and 0.0020 (NOI) in column (1), and 0.0013 (HHI) and 0.0038 (NOI) in column (2)

by close to 12.5 percentage points in both specifications. The estimates in columns (1) and (2) for $N\hat{O}I_{ijt}$ indicate a one billion dollar increase in airline j's net operating income in the year of damaging strike i decreases their repair cost disclosure probability *ex-post* the event by approximately 12 and 13 percentage points, respectively. Broadly, these results articulate airlines' voluntary repair cost disclosure is decreasing with respect to both market competitiveness and profitability.

1.5.1 Competition-profit interaction

Table 1.4 contains coefficient estimates of (1.1), including a competition-profit interaction term, again withholding and including the lagged income measure. Due to the large number of parameters included in the empirical model, estimating meaningful margins for interpreting such an effect using 2SLS can be problematic. Subsequently, estimates presented in this section are obtained using OLS, at the expense of the simultaneity concern expressed in the previous section. This was done to provide a more intuitive lens into the relationship between competition and profits on disclosure behavior.

Similar to the results found by Dedman and Lennox (2009), and following Hypothesis 1.3, the parameter estimate linked to the interaction term is positive and statistically insignificant in both specifications. Again, a one-tailed test with the null hypothesis being the parameter estimate for the interaction term is less than or equal to zero is used to determine reported significance levels, and the coefficient estimate for HHI is scaled to represent a 0.5 unit increase. The intuitive interpretation of these results is that as industry competitiveness falls, the negative effect of income on wildlife-strike-related repair cost disclosure is weakened. Specifically, profit appears to matter less to airline cost disclosure decisions under industry settings of lower competition. Figure 1.4 shows how predicted probabilities of repair cost disclosure vary across different levels of HHI, using the specification from column (1) of Table 1.4. The downward sloping lines for each level of HHI represent the negative relationship between repair cost disclosure and profitabil-

Table 1.4: OLS coefficient estimates: competition-profit interaction

	(1) $ReportCost_{ijt}$	(2) $ReportCost_{ijt}$
$NOI_{ijt} * HHI_{it}$	0.010	0.013
	(0.005)	(0.005)
HHI_{it}	-0.017	-0.025
	(0.020)	(0.020)
NOI_{ijt}	-0.114	-0.147
	(0.063)	(0.066)
NOI_{ijt-1}		-0.024
		(0.009)
Strike specific controls	Yes	Yes
Month/year FE	Yes	Yes
Airline FE/operations controls	Yes	Yes
Airport FE	Yes	Yes
N	1476	1476

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

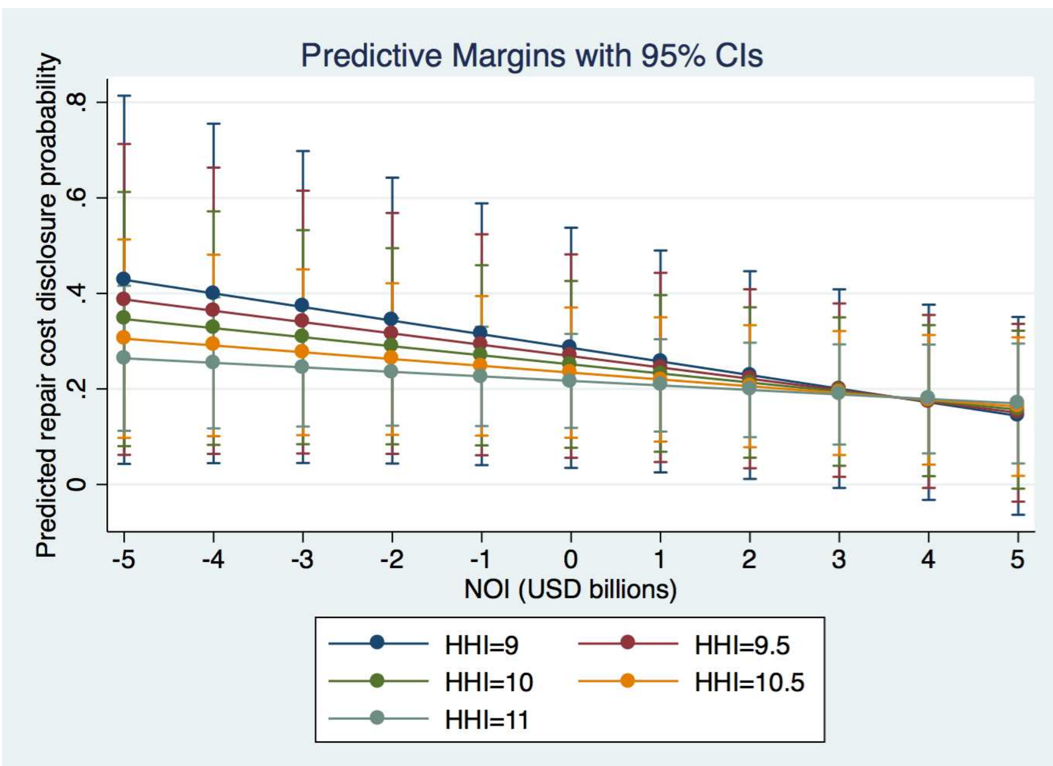
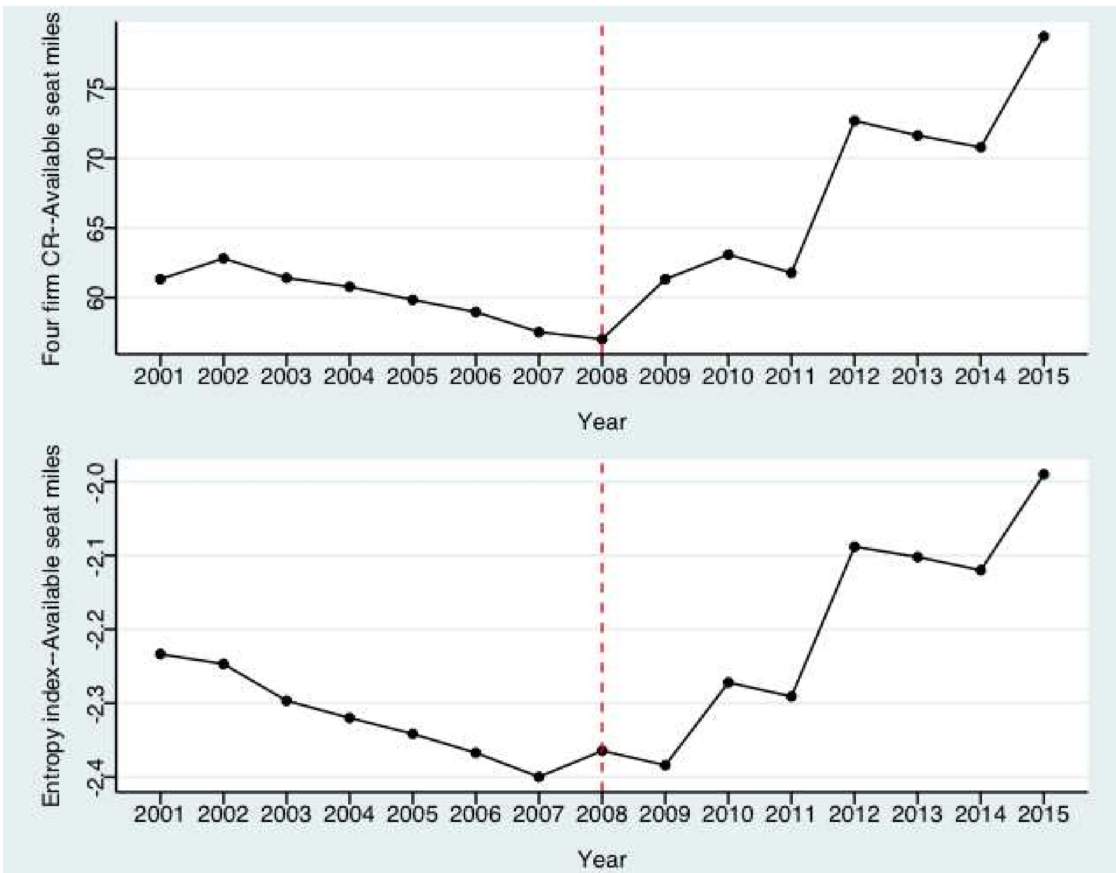


Figure 1.4: Predicted probability of repair cost disclosure for NOI across HHI

ity. The variation in the slope of each HHI value represents the interaction effect between HHI and NOI. Starting at $HHI = 9$, each subsequent 0.5 unit increase in HHI results in a flattening of the slope of the line (i.e., as market competitiveness decreases, HHI increases), the negative effect of profit on disclosure probability tapers.

1.5.2 Robustness check



Source: MIT Airline Data Project–Domestic Available Seat Miles

Figure 1.5: US domestic airline market share–four firm CR, entropy index; 2001-2015

As a robustness check of the main results, two different measures of market concentration are substituted for the HHI_{it} variable in the estimation of (1.1): the four-firm concentration ratio ($FourFirm_{it}$) and entropy index ($Entropy_{it}$). Harris (1998) notes that if the firms of interest "operate in diverse industries," using the HHI may lead to a "noisy estimate" of market concentration, and using the four-firm concentration ratio may be a preferable approach (Harris, 1998, p.125). Both indices are calculated using the same ASM data as used for HHI_{it}, and relevant summary statistics are presented in Appendix B Table B.1. Figure 1.5 plots each respective measure over the sample time period, with the red dotted line representing the Delta-Northwest merger in 2008. The overall trend of each measure mirrors that of the HHI_{it} variable, increasing significantly after the aforementioned merger event.

Table 1.5: IV average partial effects–alternate measures of concentration

	(1)	(2)	(3)	(4)
	$ReportCost_{ijt}$	$ReportCost_{ijt}$	$ReportCost_{ijt}$	$ReportCost_{ijt}$
$FourFirm_{it}$	0.0568***	0.0580***		
	(0.0180)	(0.0193)		
$Entropy_{it}$			0.468***	0.478***
			(0.148)	(0.159)
$\hat{NOI}_{ijt}$	-0.121***	-0.127***	-0.121***	-0.127***
	(0.0422)	(0.0477)	(0.0422)	(0.0477)
NOI_{ijt-1}		0.0106		0.0106
		(0.0135)		(0.0135)
Strike-specific controls	Yes	Yes	Yes	Yes
Month/year FE	Yes	Yes	Yes	Yes
Airline FE/operations controls	Yes	Yes	Yes	Yes
Airport FE	Yes	Yes	Yes	Yes
N	1476	1476	1476	1476
First stage F-stats:				
$LoadFactor_{ijt}$	30.70	25.13	30.70	25.13

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

$Entropy_{it}$ coefficient scaled by its standard deviation

Tirole (1988) defines each respective measure:

$$FourFirm = \sum_{i=1}^{4} \alpha_i \tag{1.3}$$

$$Entropy = \sum_{i=1}^{n} \alpha_i \ln(\alpha_i) \tag{1.4}$$

Where α_i again represents the market share of firm i operating in an industry with n competitors, such that $\alpha_i = q_i/Q$, $\sum_{i=1}^{n} \alpha_i = 1$, and firm market shares are ordered ($\alpha_1 \geq ... \geq \alpha_4 \geq ... \geq \alpha_n$). The four firm concentration ratio is the sum of the four highest industry market shares, while the entropy index is the sum of all market shares ($i = 1, ..., n$) times their logarithm.

Table 1.5 contains average partial effects from estimations of (1.1), substituting the variables $FourFirm_{it}$ and $Entropy_{it}$[17] for HHI_{it}, both including and excluding the one year profit lag. The significance levels reported in Table 1.5 are representative of the correct one-tailed hypotheses test corresponding to both alternate measures of concentration and NOI. The coefficients for each alternate measure are positive and statistically significant at the one percent level for all presented specifications. Furthermore, the estimated coefficient linked to net operating income retains its sign, significance, and relative magnitude across all estimations. The primary conclusion emerging from these results is that the negative relationship between competition, as measured by concentration, and voluntary airline cost disclosure appears robust to several different measures of industry concentration.

1.6 Conclusions and future research

These empirical results support existing findings regarding firm voluntary disclosure behavior, market competitiveness, and profitability. Theory suggests an inverse relationship between market competition and disclosure incentives (Arya & Mittendorf, 2007; Board, 2009; Clinch & Verrecchia, 1997; Darrough, 1993; Verrecchia, 1983). Mirroring the

[17] For a meaningful interpretation of its marginal effect, $Entropy_{it}$ is scaled by its standard deviation.

findings of Dedman and Lennox (2009) and Verrecchia and Weber (2006), the presented empirical results support this prediction through an unusually revealing context relative to the previous work. Additionally, Dedman and Lennox (2009) provide both a theoretical link and supporting empirical evidence regarding disclosure and profits such that more profitable entities reveal less. The results further align with their work in this regard.

In a corollary to the noted link between disclosure, competition, and profits, there should be a positive interaction between competition and profitability in terms of disclosure. Specifically, in the current context of decreasing industry competition, the effect of profits on the incentive to disclose is lower, although not significantly.

The results can be broken down into three fundamental conclusions. First, the recent decrease in competition appears to have reduced the competitive costs linked to wildlife-strike repair cost disclosure by US domestic airlines. Second, the most profitable of these airlines seek to hide their comparative advantage from competitors via limiting this disclosure. Finally, for a given level of market competition the most profitable should disclose the least.

This information is beneficial to decision makers on multiple levels. In terms of regulatory oversight, relevant agencies can proactively target expected non-compliers. Relatedly, such enforcement can advance the end goal of increasing the amount of knowledge available related to the economic costs of wildlife strikes. This understanding could significantly enhance airline safety, helping shape airport-specific wildlife management policies to minimize risks associated with wildlife strikes. Looking forward, the noted trend of increasing concentration in the US domestic airline market suggests that the relative increase in cost disclosure will continue under the current regulatory regime.

Future empirical research in this vein should seek out related market contexts in which the disclosure behavior of interest carries competitive costs but is unrelated to capital markets, which is a less understood indirect link between key firm incentives. Moreover, as the US domestic airline market continues to evolve in terms of market concentration, reexamining the core questions of this chapter at a future date from the vantage point of a longer time series could be revealing in itself, as well as in comparison to the present findings.

Chapter 2

Airline voluntary cost disclosure on the intensive margin: An application of Benford's Law

2.1 Introduction

This chapter of the dissertation proposes a similar question as chapter 1: Do industry competitiveness and firm-level profitability influence the *accuracy* of US domestic airlines' voluntarily disclosure of information related to repair costs resulting from a damaging wildlife strike event? The voluntary disclosure literature most relevant to this question is primarily concerned with disclosure on an *extensive* margin and its relationship to firm-level characteristics, such as profitability (e.g., Dedman and Lennox (2009)). Offering a novel empirical lens to investigate the research question are emerging methods in the economics and accounting literature that test the accuracy of self-reported data based on a statistical property exhibited by large datasets, known as Benford's Law (de Marchi & Hamilton, 2006; Dumas & Devine, 2000; Nigrini, 1996; Zahran et al., 2014). Specifically, using Benford's Law, this research tests for differences in the accuracy of voluntarily reported wildlife-strike repair costs to the NWSD across large commercial airlines categorized by both the market competition setting and relative profitability when the strike occurred.

The remainder of this chapter is organized as follows. Section 2.2 reviews the literature related to both Benford's Law and relevant voluntary disclosure research. Section 2.3 describes the data used. Section 2.4 details the empirical methods used to address the research question while offering *a-priori* hypotheses. Section 2.5 presents the results, and Section 2.6 offers a brief discussion and concluding remarks.

2.2 Literature review

2.2.1 Benford's Law

Benford's Law derives its name from physicist Frank Benford, who observed the relative frequency of first significant digits for a variety of large datasets, including city size and stock returns, to be consistent with a distribution that is uniform on a logarithmic scale (Benford, 1938). If a particular dataset can be assumed, under certain criteria, to adhere to a Benford's distribution, examining its deviation from the given expectation can be used to assess the reasonableness or accuracy of the information. It should be noted that Benford's Law is used in the context of real-world datasets like those originally examined in Benford (1938), which consist of positive numbers only. Illustrated in Figure 2.1, under

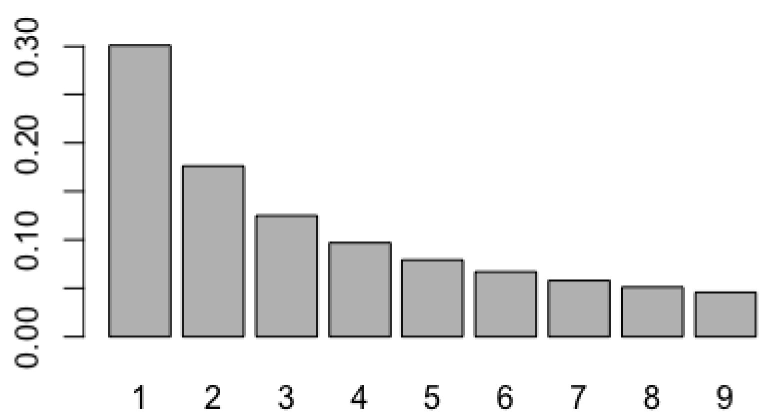


Figure 2.1: A Benford's distribution of first digits

a Benford's distribution, it is more likely that the first digit in each number of the set will be smaller than larger – e.g., a '1' versus a '9'. The relative frequencies ($f(p)$) of the first digits of such a distribution are given by the following equation[18]:

$$f(p) = log_{10}[(p+1)], p = 1, 2, ..., 9 \tag{2.1}$$

Hill (1995)'s 'General Significant Digit Law'[19] provides a generalization to N significant digits, allowing for the examination of the distribution of the first two, first three, and subsequent significant digits as a way of more rigorously evaluating data accuracy (e.g. Dumas and Devine (2000), Zahran et al. (2014))

A variety of statistical tests can be used to check for conformity to a Benford's distribution. Generally, this involves comparing the distribution of first (or N) digits of the dataset of interest to the expected distribution of first digits given by equation 2.1, or the General Significant-Digit Law. Common tests for such goodness-of-fit include Pearson's chi-square (de Marchi & Hamilton, 2006; Zahran et al., 2014), Z-scores (Dumas & Devine, 2000; Nigrini, 1996), Kolmogorov-Smirnov (de Marchi & Hamilton, 2006), and Euclidean distance (Zahran et al., 2014). Visual inspection of relevant histograms is also commonly used as a tool for conformity assessment (de Marchi & Hamilton, 2006; Dumas & Devine, 2000; Nigrini, 1996).

Varian (1972) proposes the use of Benford's Law to evaluate the performance of economic forecasting models and their input sources on the basis of realism. Using both input data and data generated from the Bay Area Simulation Study (BASS) IV model, Varian compares the distribution of first digits provided by the model and the input data to that predicted by Benford's Law, finding both datasets conformed relatively well to the Benford distribution.

[18]This is a formal statement of Benford's Law (Dumas & Devine, 2000).

[19]Dumas and Devine (2000) replicate Hill's derivation by showing $Prob(\cap_{i=1}^{k} D_i = d_i) = log_{10}[1 + (\sum_{i=1}^{k} d_i * 10^{k-i})^{-1}]$ where D_i is the i^{th} significant digit of x, $k \in$ natural numbers, the first significant digit $d_1 \in 1, 2, ..., 9$, for $j = 2, ..., k$.

Dumas and Devine (2000) introduce Benford's Law as a way to detect non-compliance in the self-reporting of emissions data, providing a broad overview of methods and tests to detect evidence of the ex-post manipulation of self-reported emissions data. The authors investigate data regarding annual volatile organic compounds (VOC) for firms in North Carolina over the 1996-1998 time period, focusing on differences in reporting behaviors between large Title V facilities and small facilities. Results from the application of Nigrini (1996)'s Distortion Factor Model (DFM)[20] indicate emissions data reported by all types of firms contain distortions that reduce the mean value of emissions approximately 9.5-10% below the expectation given by Benford's Law. Additionally, the relative distortion of reported emissions between Title V and small facilities was not significantly different, indicating similar self-reporting behaviors between the two firm types.

de Marchi and Hamilton (2006) employ Benford's Law to assess the accuracy of self-reported pollution figures from plants emitting chemicals that are included in the Toxic Release Inventory (TRI). Their primary finding is that at least two out of twelve chemicals for which emissions are self-reported by TRI firms appear to be misrepresentative of their actual emissions levels based on the relatively good fit of EPA monitor data for the same two chemicals (lead and nitric acid) to a Benford's distribution.

Zahran et al. (2014) use a Benford's distribution to test whether the EPA's Final Rule policy (implemented in 2001) improved the accuracy of firms' self-reported lead emissions levels.[21] The authors build on the work of de Marchi and Hamilton (2006) by incorporating quasi-experimental difference-in-difference techniques to test for changes in reporting accuracy over time, validating Benford's Law as an analytic tool by comparing self-reported emissions across various levels of reporting discretion (i.e., fugitive, stack, and off-site transfers). They additionally extend their analyses to include the second and third digit tests to examine the conformity of self-reported data to a Benford's distribution. The anal-

[20]The Distortion Factor Model allows for testing the magnitude and direction of bias if such evidence is found in a given dataset (Dumas & Devine, 2000; Nigrini, 1996).

[21]This was one of the chemicals de Marchi and Hamilton (2006) found to be reported inaccurately.

ysis concludes that firm-reported emissions data increased significantly in accuracy after the Final Rule Policy was implemented.

2.2.2 Market competition, profitability, and disclosure accuracy

As described in Section 1.2 of Chapter 1, most of the voluntary disclosure literature, both theoretical and empirical, proposes that on the *extensive* margin, the relationship between firm disclosure decisions and both market competitiveness (Arya & Mittendorf, 2007; Board, 2009; Clinch & Verrecchia, 1997; Darrough, 1993; Verrecchia & Weber, 2006) and profitability (Dedman & Lennox, 2009) is negative. Specifically, the costs of voluntary disclosure are hypothesized to be greater for more profitable firms and in relatively more competitive market settings. Moreover, these propositions appear to hold robustly when examining firm decisions outside the realm of capital markets (e.g., Dedman and Lennox (2009), Chapter 1) as done in this and the previous chapter.

Translating these arguments to the nexus of disclosure accuracy, it can be inferred that if the costs of disclosure on the *intensive* margin are increasing via the same mechanisms impacting disclosure on the *extensive* margin, disclosure accuracy should be decreasing with respect to both market competition and firm-level profitability. More simply stated, if the cost of disclosing information accurately is higher under more competitive product market conditions, as well as for more profitable firms in the same fashion as the binary decision to disclose, voluntary disclosure should be less accurate for such competition conditions and firms.

The literature is sparse regarding the use of Benford's Law empirically as a way to examine the impact of market forces on disclosure. A single study, Rauch, Goettsche, Mouaaouy, and Geidel (2013), applies Benford's Law to self-reported price data by Western Australian petroleum producers. The authors find deviation from a Benford's distribution of first digits to be significantly correlated with 'well-known' firm-level indicators of anti-competitive behavior.

2.3 Data

The dataset used in this chapter consists of 494 nominal repair costs (USD) voluntarily reported by US domestic air carriers to the NWSD after a damaging wildlife strike event during the time period 2000-2015. Each observation contains a reporting airline to the NWSD. Additional utilized firm-level data was obtained from MIT's Airline Data Project. Airlines that merged with another airline do not appear in the sample after the year the merger was finalized. This sample contains all observations from the dataset employed in Chapter 1 that disclosed a repair cost, plus additional cost-disclosure observations that were omitted from the first chapter due to insufficient information about the associated wildlife strike. Table 2.1 presents relevant summary statistics, while Figure 2.2 shows a

Table 2.1: Summary statistics: all reported costs

Statistic	Reported repair cost (USD)
N	494
Mean	248,330.40
St. Dev.	923,919.60
Min	25
Median	26,150
Max	14,000,000
Skewness	9.600
Kurtosis	121.792

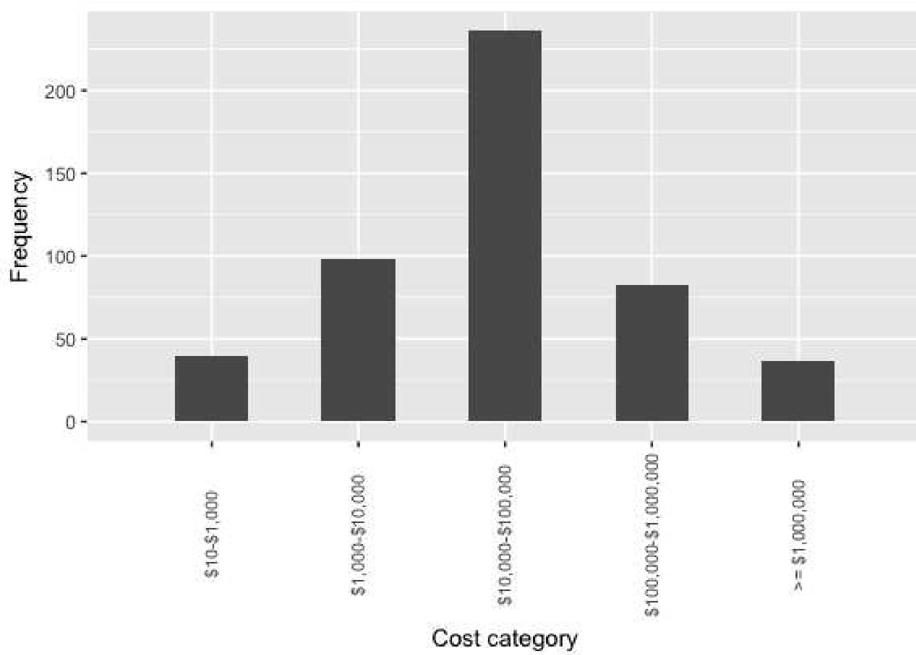
categorical distribution of reported sample costs. The average reported cost in the sample is approximately $248,330, with a maximum repair cost of $14,000,000.

Dumas and Devine (2000)[22] provide several statistical 'rules of thumb' to assume a given dataset conforms to a Benford's distribution:

1. *A single type of phenomena*
2. *Several orders of magnitude*
3. *No theoretical minimum or maximum, excluding zero*

[22]See also Cho and Gaines (2007); Durtschi, Hillison, and Pacini (2004).

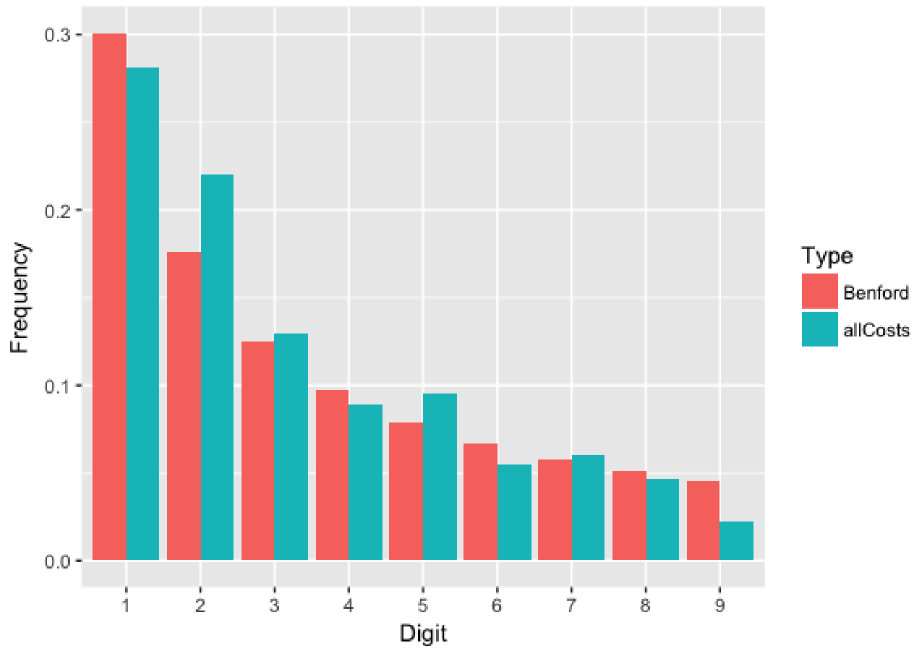
4. *Positive skewness (i.e., a higher frequency of relatively small numbers than large numbers)*
5. *Mean of the data larger than its median*
6. *No systematically duplicated or assigned numbers, such as account numbers or user IDs*



Source: NWSD

Figure 2.2: Sample distribution of reported repair costs; 2000-2015

Examining Table 2.1 and Figure 2.2, the dataset appears to adhere to the above criteria. Table 2.1 shows the single phenomenon (i.e., wildlife-strike-damage costs) to cover several orders of magnitude, while also possessing a median greater than its mean and positive skewness. Figure 2.2 confirms this skewness; most of the costs are less than \$100,000. Figure 2.3 presents the first significant digit distribution of the sample to the distribution we would expect under Benford's Law. The sample conforms relatively well. Two goodness-of-fit statistics, Pearson's chi-square and Morrow (2010)'s modified Euclidean distance measure indicate no significant deviation from a Benford's distribution using a one percent significance level.



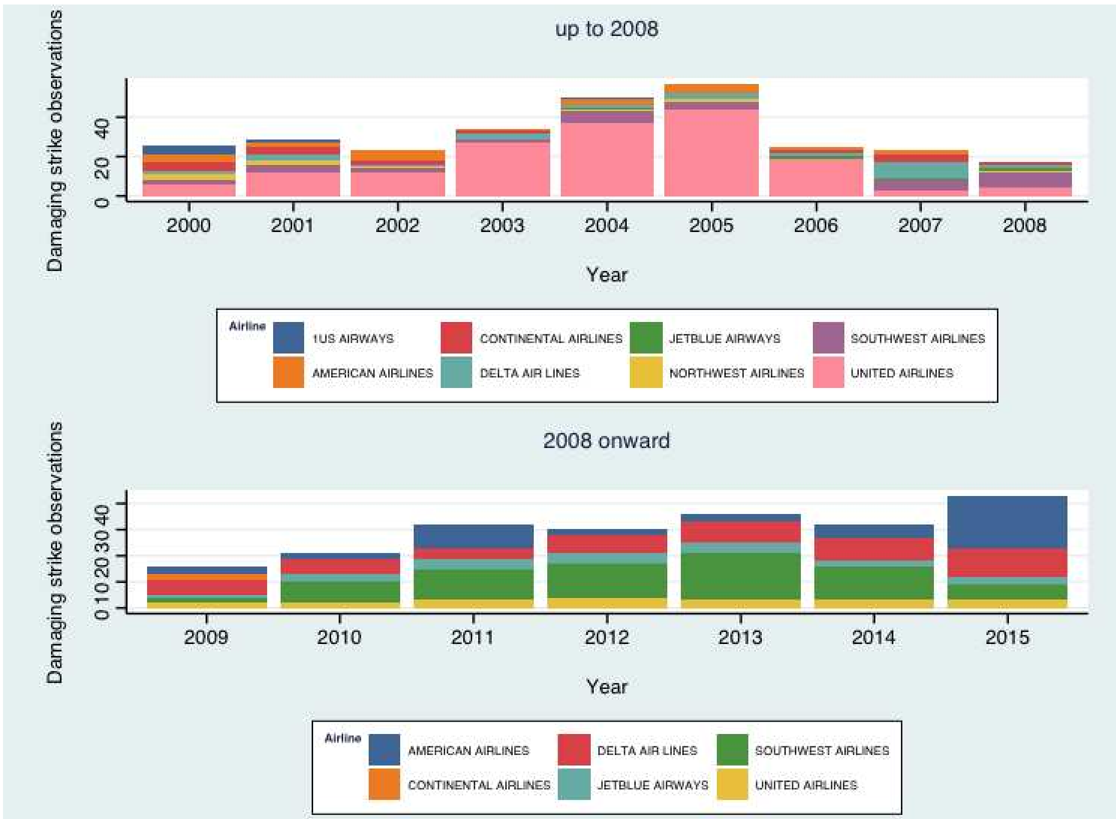
Source: NWSD

Figure 2.3: Distribution of first digits: All reported costs

To examine the relationship between market competition, profitability, and disclosure accuracy, observations from the sample are categorized into separate groups for each disclosure determinant of interest (i.e., competition and profitability) to ultimately test for differences in reporting behaviors across assigned groups. The following subsections describe this partitioning process in detail.

2.3.1 Market competition

The combined market share of the top four domestic US airlines increased from 65% in 2010 to 84% in 2015. This is largely the result of several consolidations within the industry, beginning with the merger of Delta Air Lines and Northwest Airlines in 2008. Subsequent mergers include United and Continental Airlines in 2010, Southwest Airlines and AirTran Airways in 2011, and American Airlines and US Airways in 2013. Several studies have linked this consolidation to a decline in industry competitiveness (Peterman, 2014; Shen, 2017).



Source: NWSD/MIT-ADP

Figure 2.4: Sample airline composition: Up to and after 2008

As an admittedly imperfect categorization of market competitiveness facing disclosing airlines at the time of their decisions, whether the cost was reported prior to 2008, the year of the Dela-Northwest merger, or afterward is differentiated. Figure 2.4 presents the distribution of observations across airlines and years for each respective category. A clear trend emerges in that the variety of airlines in the up to 2008 sample is greater. Following Chapter 1, this analysis uses theoretical intuition from the Cournot oligopoly model. In this model, as the number of firms in the market increases, the market price trends toward a perfectly competitive price (Tirole, 1988). The following logic is that because of the greater number of competing airlines in the up to 2008 period, firms in this period faced a relatively more competitive product market setting. Studies including Brander and Zhang (1990) and Desgranges and Gauthier (2016) provide support for assuming that competition in the US domestic airline industry is consistent with the Cournot model.

Tables 2.2 and 2.3 include summary statistics for each respective grouping. The up to 2008 period contains 284 observations with a mean reported cost of $297,717, while the 2008 onward period contains 210 reported costs with a similar mean of approximately $286,989. Similar to the whole sample, each separate dataset appears to meet the criteria necessary to assume adherence to Benford's Law, including several orders of magnitude, a mean larger than the median, and positive skewness. Both groups' costs distributions have a minimum less than $100 and a maximum of over $8,000,000. The 'up to 2008' group's median reported cost is $25,000, smaller than the foregoing average. Similarly, the '2008 and after' group's median of $37,168 is less than its previously reported mean. Skewness is close to 8 and 9 for the up to 2008 and 2008 onward groups, respectively. While the size of each category is less than ideal, subsequent hypothesis testing attempts to address this through the use of a modified test-statistic (Morrow, 2010) designed specifically for implementation with small samples and Benford's Law.

Table 2.2: Summary statistics: up to 2008

Statistic	Reported repair cost (USD)	Net operating income (USD billions)
N	284	284
Mean	231,322.50	-0.641
St. Dev.	695,220.10	1.508
Min	70	-8.314
Median	25,000	-0.219
Max	8,925,119	1.637
Skewness	7.839	
Kurtosis	89.498	

Table 2.3: Summary statistics: 2008 onward

Statistic	Reported repair cost (USD)	Net operating income (USD billions)
N	210	210
Mean	271,331.60	2.136
St. Dev.	1,165,415.00	2.339
Min	25	-1.170
Median	37,168	1.278
Max	14,000,000	7.802
Skewness	9.005	
Kurtosis	97.846	

Summary statistics for airline net operating income in USD billions in the year of the reported strike event are also provided in Tables 2.2 and 2.3. Unsurprisingly, and indicative of the variation in product market competitiveness across categories, the average airline operating income in the 2008 onward period is substantially higher than the up to 2008 period and further possess a maximum reported income $6 billion greater than the highest airline income for the up to 2008 period.

2.3.2 Profitability

To test for differences in disclosure accuracy between airlines of different categories of profitability, observations are divided based on whether the reporting airline was either below or at/above the sample average net operating income of all airlines for the year of the reported strike. This method was primarily chosen to capture the *relative* profitability conditions a given airline was facing at the time of reporting.

Figure 2.5 presents the sample composition of airlines across time for each respective category (i.e., below, above). Note that two observations with the same corresponding airline fall into separate categories conditional on the year the strike was reported. For example, a cost reported by American Airlines in the year 2000 would fall into the below category because they were below the sample average for net income in the same year. Conversely, another cost reported by American Airlines in the year 2015 would be

Source: NWSD/MIT-ADP

Figure 2.5: Sample airline composition: Above/below average profitability

placed in the above category because American Airlines possessed a net operating income above the sample average of reporting firms for 2015. Essentially, if the repair cost was reported by an airline that was below the sample average on the profitability metric in the year of the reported event, it is categorized as below. If the observed cost is reported by an airline that was at or above the sample average profitability for the year of the strike event, it is classified as above. For a given year, a cost reported by a specific airline can only appear in a single category.

Tables 2.4 and 2.5 display summary statistics for each profitability category. The below group has a sample size of 278 and a mean reported cost of approximately $180,903. The sample size for the above group is 216, and its associated average reported cost is close to $335,113. Each partitioned sample again posses the first-pass criteria for effective use

Table 2.4: Summary statistics: below average profitability

Statistic	Reported repair cost (USD)	Net operating income (USD billions)
N	278	278
Mean	180,902.50	-0.193
St. Dev.	561,614.30	1.968
Min	25	-8.314
Median	26,000	0.322
Max	6,500,000	5.166
Skewness	6.900	
Kurtosis	65.701	

Table 2.5: Summary statistics: above average profitability

Statistic	Reported repair cost (USD)	Net operating income (USD billions)
N	216	216
Mean	335,112.600	1.483
St. Dev.	1,239,990.00	2.463
Min	71	-1.683
Median	27,150	0.458
Max	14,000,000	7.802
Skewness	8.124	
Kurtosis	80.604	

of Benford's Law. Both the below and above groups' medians of \$26,000 and \$27,150 are below their respective means. Both groups also have positive skewness. The below group has a skewness of close to 7, while the above group's is approximately 8. Each group's cost distribution spans the necessary several orders of magnitude. The below group's minimum reported cost is \$25, while its maximum is \$6,500,000. Similarly, the above group's minimum is \$71, and its maximum is \$14,000,000

The following section describes the empirical methods used to test for differences in reporting accuracy between each category related to both market competition and profitability while also formally stating the specific hypotheses tested.

2.4 Methods and hypotheses

2.4.1 Methods

Two goodness-of-fit statistics commonly used to calculate the extent to which observed first digit distributions in reported data deviate from a Benford's distribution are utilized to examine the relationship between market competition, firm-level profits, and the accuracy of voluntarily-reported direct wildlife-strike-damage costs by US domestic airlines. Similar to Zahran et al. (2014), the first statistic used is Pearson's chi-square, given by the following equation:

$$\chi^2 = N \sum_{i=1}^{9} \frac{(f_i^o - f_i^e)^2}{f_i^e} \tag{2.2}$$

Where f_i^o and f_i^e represent the observed and expected frequencies given by a Benford's distribution for the ith digit, respectively. This statistic is calculated with 8 corresponding degrees of freedom (9-1=8), and the associated null hypothesis is conformity to a Benford's distribution. However, Morrow (2010) notes the limitations of the chi-square test on small samples, stating "due to its low power for even moderately small sample sizes, it is often unsuitable." (Morrow, 2010, p.3). Moreover, test variations similar to the chi-square goodness-of-fit, including the Kolmogorov-Smirinov and Kuiper tests, have been shown to be overly conservative when used in conjunction with Benford's Law due to underlying assumptions attached to the observed distribution of first digits (Morrow, 2010). In this context, Morrow (2010) introduces a modified Euclidean distance statistic specifically for use with Benford's Law, formally expressed as follows:

$$d_N^* = \sqrt{N} * \sqrt{\sum_{i=1}^{9} (f_i^o - f_i^e)^2} \tag{2.3}$$

This statistic is used in the analysis to account for potential sample size limitations, formally shown in Morrow (2010) to perform sufficiently well for samples such that 50 $\leq N \leq$ 500.

Both of the above statistics test for the *presence* of bias in a given dataset. Of supplemental interest to this investigation is the direction of the bias, if present, and whether this relative bias differs between the categories of market competitiveness and profitability. Nigrini (1996)'s Distortion Factor Model (DFM) provides a way to answer this question by comparing the observed mean of a collapsed version of data to the expected mean of a dataset with the same number of observations, distributed following Benford's Law (Dumas & Devine, 2000; Nigrini, 1996). Nigrini (1996, p.76) describes this process in detail, listing the following steps used to calculate relative distortion for a particular dataset:

1. *Transform reported numbers to numbers in the range (10, 100).*
   - (a) *Delete all numbers that are less than 10, including all numbers reported as zero dollars. This step ensures that all numbers have an explicit first and second digit.*[23]*.*
   - (b) *Collapse all reported numbers equal to or greater than 100 to the range (10,100) by moving the decimal point as required.*
2. *Compute the actual mean (AM) of the collapsed numbers.*
3. *Compute the expected mean (EM) of the observations of a Benford Set scaled to the (10,100) range using the following equation:*

$$EM = \frac{90}{[N * (10^{1/N} - 1)]} \tag{2.4}$$

4. *Compute the distortion factor (DF):*

$$DF = \frac{(AM - EM)}{EM} \tag{2.5}$$

Expressed as a percentage, the DF measures the percent deviation of the AM from the EM. When the DF takes on a negative value, it indicates that more smaller numbers were

[23]Note that the sample used in this chapter has no reported costs less than \$10; see Table 2.1.

being used than expected from a Benford set of the same size. A positive value attached to the DF implies an excessive use of large digits relative to the expectation offered by Benford's Law. The following subsection articulates the specific hypotheses tested in this chapter using the three aforementioned test statistics.

2.4.2 Hypotheses

Market competition

As discussed in Section 2.3, observations are grouped into two different categories based on when the damaging strike occurred: the relatively more competitive setting prior to the 2008 Delta-Northwest merger (*UpTo-2008*) and the less competitive setting, which includes strikes that occurred in the year 2008 and onward (*2008-Onward*). Separate hypotheses will be tested using the methods detailed in Section 2.4 for each market competitiveness grouping.

The first method tests for the presence of bias in each category by using chi-square (χ^2) and modified Euclidean distance statistics (d_N^*) to examine the relationship between the first digit distribution of reported repair costs. This analysis formally tests the following hypothesis related to repair cost disclosure accuracy and market competition:

- *Hypothesis 2.1.A:*
  - H_0: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines prior to 2008 *is not* different from a Benford's distribution of first significant digits.
  - H_A: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines prior to 2008 *is* different from a Benford's distribution of first significant digits.
- *Hypothesis 2.1.B:*

- H_0: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines in 2008 and following *is not* different from a Benford's distribution of first significant digits.
- H_A: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines in 2008 and following *is* different from a Benford's distribution of first significant digits.

In light of the literature review in Section 2.2, as well as Chapter 1 and Section 1.2, the before-the-fact expectation is that under more competitive product market settings (*up to 2008*), costs reported in this period are relatively more probable to be *less* accurate or, alternatively, display a significant deviation from the Benford's distribution. However, under the relatively less competitive 2008 and onward market scenario, reported costs can be expected to display relatively more accuracy and, thus, are more likely to conform to Benford's Law. Combining the two hypotheses informally, the expected case is $\chi^2_{UpTo-2008} \geq \chi^2_{2008-Onward}$ and $d^*_{N,UpTo-2008} \geq d^*_{N,2008-Onward}$.

The other formally tested hypothesis relates to the distortion factor (DF), testing whether relative bias is different between the two classifications of market competitiveness. This hypothesis is stated formally as follows:

- *Hypothesis 2.2*:
  - H_0: $DF_{UpTo-2008} = DF_{2008-Onward}$
  - H_A: $DF_{UpTo-2008} \neq DF_{2008-Onward}$

Following Dumas and Devine (2000), this is specifically performed using a Z-test of difference in means where

$$Z = \frac{(DF_{UpTo-2008} - DF_{2008-Onward})}{\sqrt{\frac{s^2_{UpTo-2008}}{N_{UpTo-2008}} + \frac{s^2_{2008-Onward}}{N_{2008-Onward}}}} \tag{2.6}$$

and $N_{[...]}$ and $s^2_{[...]}$ represent the sample size and DF variance for each respective category. Assuming competition varies sufficiently across categories, relevant theory and empirical results imply rejecting the null hypothesis, such that the relative bias present in reported costs by each group differ from each other.

Profitability

Like the partitioning related to market competition, observations are grouped into separate categories based on relative industry profitability during the year of the recorded strike event. The hypotheses related to the profit-disclosure accuracy nexus, which tests for the presence of bias (χ^2, d_N^*) through first digit analysis similar to Hypotheses 2.1.A and 2.1.B are as follows:

- *Hypothesis 2.3.A:*
  - H_0: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines with profits below the industry average profit for the year of the recorded strike event *is not* different from a Benford's distribution of first significant digits.
  - H_A: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines with profits below the industry average profit for the year of the recorded strike event *is* different from a Benford's distribution of first significant digits.
- *Hypothesis 2.3.B:*
  - H_0: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines with profits at or above the industry average profit for the year of the recorded strike event *is not* different from a Benford's distribution of first significant digits.
  - H_A: The distribution of first significant digits of direct repair costs related to wildlife strikes voluntarily reported by US domestic airlines with profits at or above the industry average profit for the year of the recorded strike event *is* different from a Benford's distribution of first significant digits.

Appealing to the literature discussed in Sections 2.2 and 1.2, the competitive costs of voluntary disclosure on the *intensive* margin are assumed to be increasing with respect to airline profitability. Subsequently, costs reported by airlines relatively more profitable in the year of the strike event who have 'more-to-lose' should be expected to be *less* accurate than those reported by less profitable airlines. In this context, it is plausible to predict a more likely rejection of the null for Hypothesis 2.3.B as opposed to Hypothesis 2.3.A. Specifically, deviation from a Benford's distribution by the first digits of damages reported by more profitable airlines, who face higher competitive costs related to voluntary repair cost disclosure than relatively less profitable firms, is assumed to be more likely. Informally combining these two hypotheses, the author expects $\chi^2_{Above} \geq \chi^2_{Below}$ and $d^*_{N,\,Above} \geq d^*_{N,\,Below}$

Given below and in the same spirit as Hypothesis 2.2, Hypothesis 2.4 employs the distortion factor and Z-test of difference in means to test if the relative bias for each profit category is different from each other:

- *Hypothesis 2.4*:
  - H_0: $DF_{Below} = DF_{Above}$
  - H_A: $DF_{Below} \neq DF_{Above}$

The following section presents results from empirical tests (using the previously discussed methods) of all the hypotheses listed in this chapter.

2.5 Results

2.5.1 Market competition

Table 2.6 presents test results related to market competition and disclosure accuracy. The associated chi-square values for the up to 2008 and 2008 onward categories are 25.731 and 17.613, respectively. Similarly, their associated modified distance statistics (d^*_N) are approximately 1.850 and 1.136, respectively. As expected, both goodness-of-fit statistics are larger for the more competitive up to 2008 group relative to costs reported in 2008 and after. Furthermore, both statistics imply that the first digit of costs reported

Table 2.6: First digit analysis results: competition

Classification	N	χ^2	P-Value$_{\chi^2}$	d_N^*	P-Value$_{d_N^*}$	DF
up to 2008	284	22.925	.003	1.850	.001	-10.67
2008 onward	210	17.613	.024	1.136	.152	-5.02

by the up to 2008 category are different using a 1 percent level of significance from a Benford's distribution of fist significant digits. The associated critical values for a 1 percent level of significance are 13.36 for Pearson's chi-square and 1.569 for Morrow (2010)'s modified distance measure, thus rejecting the null of Hypothesis 2.1.A.

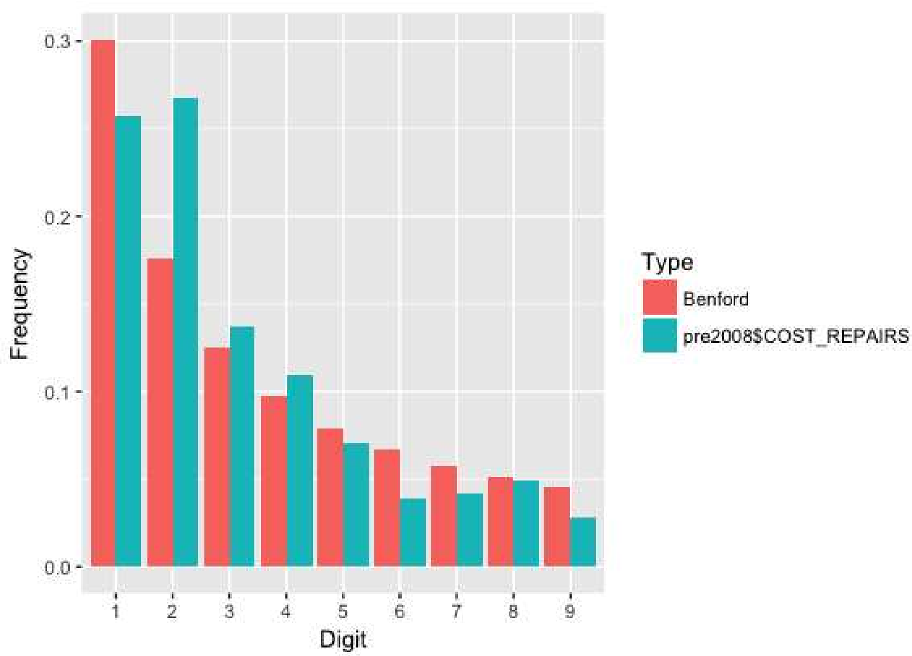


Figure 2.6: Distribution of first digits: Up to 2008

Figure 2.6 presents a histogram of the frequency of first significant digits for the up to 2008 category and Benford's Law. An obvious spike is present for the digits 2, 3, and

4, while a valley exists for the digits 6, 7, and 9. This implicates an excess of small first digits and a lack of large first digits relative to expectations offered by Benford's Law. This is confirmed by the groups' distortion factor measure of -10.67, indicating its average reported costs are 10.34% less than the expected mean of a similar distribution that follows Benford.

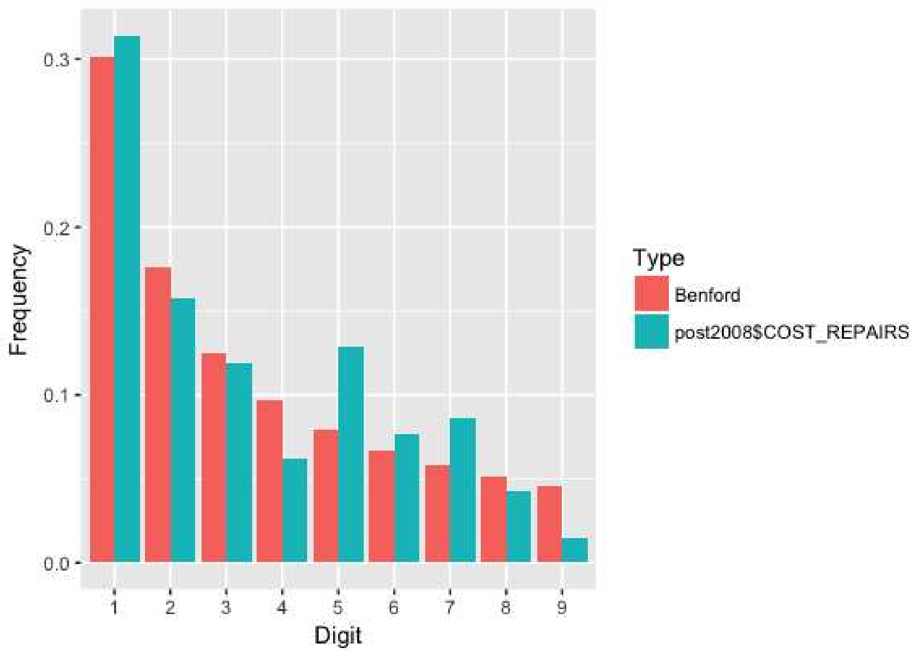


Figure 2.7: Distribution of first digits: 2008 onward

For Hypothesis 2.1.B and the 2008 onward category, test results show a failure to reject the associated null using a one percent level of significance for both the chi-square statistic and the small sample-specific distance metric.[24] The group's distortion factor measure is approximately -5.02, again signaling a negative bias attached to their reported costs. This bias can be seen in Figure 2.7 where relative spikes are present for the first digits 1 and 5 compared to larger digits, indicating the over-reporting of costs possessing these first significant digits.

[24]Using the same critical values of 13.36 and 1.569 for chi-square and Morrow (2010)'s distance measure, respectively.

Results from the Z-test of difference-in-means for Hypothesis 2.2 indicate that the relative bias (distortion factor) in reported costs differs between groups. The calculated Z-statistic using (2.6) is 14.982, which for a critical value of 2.576, rejects the related null hypothesis at a one percent significance level and concludes the relative bias present in reported costs is different from costs reported to the NWSD by US domestic airlines prior to 2008 and those reported in and after 2008.

2.5.2 Profitability

Results related to profitability are presented in Table 2.7. For the below category, their chi-square and d_N^* values are 7.046 and 0.956, respectively. Testing Hypothesis 2.3.A, again using a one percent level of significance indicates no significant deviation from a Benford's distribution, which is a failure to reject the associated null.

Table 2.7: First digit analysis results: profitability

Classification	N	χ^2	P-Value$_{\chi^2}$	d_N^*	P-Value$_{d_N^*}$	DF
Below industry Average profit	278	7.046	.532	0.956	0.346	-5.41
Above industry Average profit	216	10.557	.229	0.984	.311	-11.950

The group's distortion factor metric is -5.41, indicating the average cost reported by this category is 5.41 percent lower than the expected average of a same-size sample with first digits that follow a Benford's distribution. This is visualized in Figure 2.8 via the large spike at 2 and the valley at 9. Regarding Hypothesis 2.3.B and the above group, their test-statistic values of $\chi^2 = 10.557$ and $d_N^* = 0.984$ fail to reject the null using a one percent level of significance. Noting the matching *a-priori* expectations, it is in fact the case: $\chi^2_{Above} \geq \chi^2_{Below}$ and $d^*_{N,\,Above} \geq d^*_{N,\,Below}$.

The category's distortion factor of -11.95 is substantially larger than the below category and indicative of negative bias. This is demonstrated in Figure 2.9 by the presence of

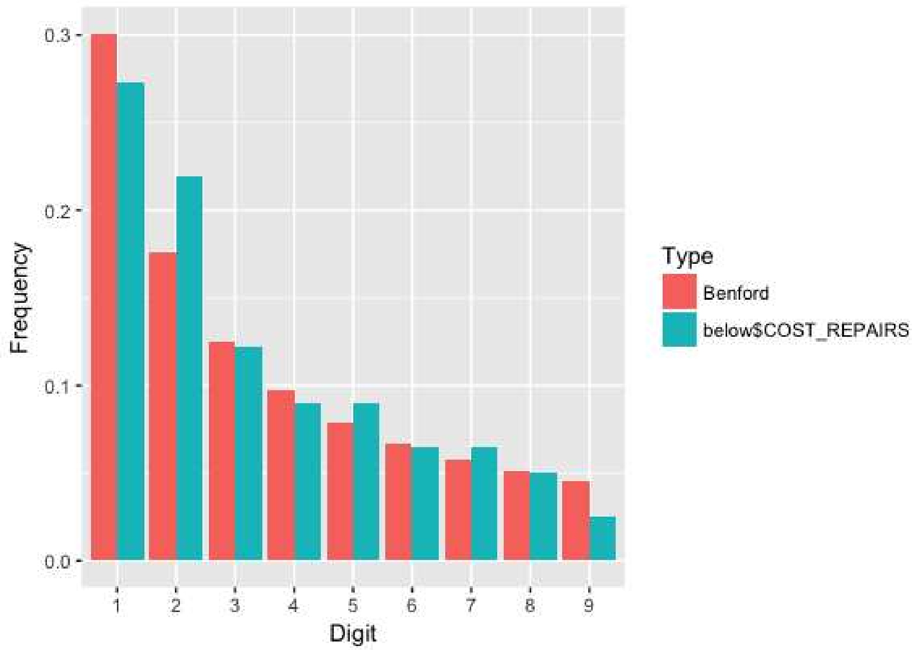


Figure 2.8: Distribution of first digits: Below average profitability

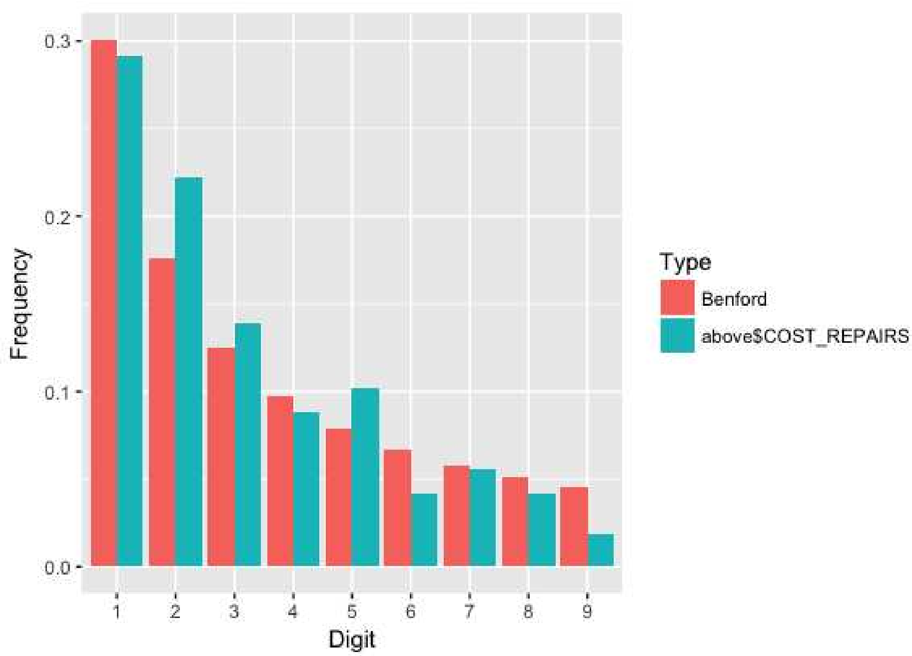


Figure 2.9: Distribution of first digits: Above average profitability

relatively more first digits of 3 and 5, accompanied by the lack of digits 6, 8, and 9. Z-test results using a calculated Z-statistic of 17.47 and a critical value (one percent level of significance) of 2.576 rejects the related null and implies that the relative bias present in costs reported by the below and above group to be different from each other.

2.6 Discussion and conclusion

Broadly, these results imply differences in the accuracy of reported direct repair costs across categories of airline-level observations classified according to market competitiveness and relative profitability at the time of a wildlife strike event.

Regarding market competition, due to the imperfect method of categorization used in this chapter, it is difficult to infer whether the differences in reporting accuracy are truly driven by variation in industry competitiveness or some other unobservable factor changing over time. It should be noted that this result generally matches theoretical predictions, such that costs of disclosure on the *intensive* margin appear lower during times of decreased market competition. One reassuring takeaway is that costs disclosed more recently appear to be more accurate than past disclosures. This is potentially indicative of improved compliance by airlines with existing regulatory conditions, which have remained generally unchanged across the sample time period.

The results regarding profitability are, perhaps, relatively weaker empirically but, similarly, match before-the-fact expectations in that the costs associated with voluntary disclosure behavior presented in this chapter appear to be increasing with respect to firm profitability. While neither the below average nor the above average group's reported costs differ from a Benford's distribution under both goodness-of-fit measures, both statistics are larger for the above industry average profit categorization than the below group. Likewise, the above group's distortion factor is different from and larger than the below category.

Due to its unique context, and similar to Chapter 1, one limitation of these findings is external validity, at least in the realm of voluntary disclosure behavior. Furthermore, this analysis has attempted to show Benford's Law can be a useful empirical tool, with the ultimate goal of broadening the related, limited line of research. Future work is warranted in terms of using Benford's Law as a way to identify behavioral responses to variation in fundamental market forces, including but not limited to competition and profit (e.g., Rauch et al. (2013)).

Chapter 3

Estimating the cost of wildlife strikes at US airports: A machine learning approach

3.1 Background, literature, motivation

Costs associated with wildlife-aircraft collisions in the US are widely acknowledged by the aviation community (Anderson et al., 2015; DeVault, Belant, Blackwell, & Seamens, 2011; Dolbeer et al., 2014). However, previous estimates of national-level *direct* costs of wildlife strikes to aircraft (costs directly linked to structural damage to the aircraft) have made simplifying assumptions difficult in the face of data limitations, which may bias estimates of the true magnitude. Information used to calculate economic damage estimates, such as Dolbeer et al. (2014)'s \$157 million estimate of direct costs from wildlife strikes annually in the US, rely on the voluntary disclosure of repair cost information resulting from a damaging strike event by aircraft operators to the NWSD. This chapter finds that of a total 14,614 strikes reported to the database that indicate structural damage to the aircraft, only 4,226 ($\approx$30%) report an associated repair cost.

Employing variations of mean cost assignment to combat this missing data problem, various estimates, such as Dolbeer et al. (2014)'s, have assumed all damaging strikes to the NWSD have similar costs. If reporting behaviors vary across airports such that those with similar levels of average strike costs are under or over reporting relative to others, aggregate measures such as Dolbeer et al. (2014)'s may understate or overstate the true economic burden of wildlife strikes at the national level. At the airport level, some may suffer above or below average damage costs and any evaluation of costs or benefits from a reduction in damaging strikes using an aggregate average is potentially biased.

Moreover, it may also be the case that damaging strikes are more likely to be reported than less serious non-damaging strikes.

In response to this concern and limiting the context specifically to direct costs from bird strikes, Anderson et al. (2015) present a two-part empirical model using the NWSD that first estimates the probability of a damaging bird strike as a function of strike-specific characteristics, including plane size and engine type, as well as the size and number birds struck. The second component of the model predicts repair costs conditional on the occurrence of a non-zero damage strike. As noted by the authors, their model is potentially useful for interpolating the repair costs of strikes in the NWSD that fail to report a damage cost. While interpolating data in this manner cannot circumvent the problem of the overall underreporting of strikes, it can mitigate the problem of systemic underreporting of direct damage costs (Anderson et al., 2015; Dolbeer, 2015; Dolbeer & Wright, 2009; Dolbeer et al., 2014).

Although statistical models, such as the one used by Anderson et al. (2015) may be the optimal choice for gaining *inference* (e.g., β coefficient estimates) related to specific factors that influence the probability and cost of a damaging wildlife strike, machine learning techniques, including classification trees, random forests, and artificial neural networks, have recently emerged as a potentially more useful method for the problem of *prediction* in economics (e.g., $\hat{y}$), which is the focus of this chapter (Cooper, 1999; Mullainathan & Spiess, 2017; Verlinden, Duflo, P.Collin, & Cattrysse, 2008). In the presence of many input variables, Mullainathan and Spiess (2017) find several machine learning algorithms, including random forest, outperform ordinary least squares regarding predictions of out-of-sample data. Verlinden et al. (2008) estimate the costs of sheet-metal production projects by feeding various variables, including material type, sheet thickness, and the number of holes through an artificial neural network in addition to using traditional econometric techniques. They show that while a cost of loss of parameter interpretation is incurred, using a neural network improved prediction performance. Relatedly, Cooper (1999) finds an artificial neural network to out-predict multiple regression in classifying countries who seek to reschedule their international debt.

The goal of this chapter is to provide a more robust measure of the economic burden of wildlife strikes in the United States by comparing traditional econometric techniques (e.g., ordinary least squares regression) with emerging machine learning methods, including a random forest and artificial neural network to develop the most accurate tool for use in interpolating missing cost data in the NWSD.

Two immediate implications of the exercise of interpolating missing observations in the database appear at the national- and airport- levels. At the national level, a comparison of a more accurate prediction-driven measure of the annual economic burden from wildlife strikes to Dolbeer et al. (2014)'s most recent estimate provides potential insight into the nature of bias, if any, resulting from the prior use of naive methods, in turn offering a new perspective regarding the magnitude of the wildlife strike problem in the US. This alternative lens will inform federal policymakers in charge of budget allocation to various wildlife management programs.

At the airport level, this research provides a potential tool for program efficiency evaluation. Typically, airports are restricted to using a specific average cost per strike metric (based on NWSD observations) recorded at their airport (or the national average if no information is available) to measure economic benefits in terms of strike reduction. This is potentially misleading if the available reported costs are not representative of the 'true' typical wildlife strike at a given airport. If airports can accurately predict an average cost based on characteristics of their observed strikes, such as the aircraft type and animal size, this will provide a more precise and individualized way to evaluate the economic benefits of wildlife strike mitigation.

Lastly, this chapter informs by applying new tools to the problem of prediction necessitated by the presence of missing information. For future researchers facing a similar task, the endeavor will provide a reference for comparing machine learning techniques to previous strategies used to deal with this general prediction issue, such as econometric modeling. The remainder of this chapter proceeds as follows. Section 3.2 provides a brief discussion of each investigated machine learning algorithm. Section 3.3 describes the data. Section 3.4 details the model selection process, while Section 3.5 interpolates missing costs in the NWSD using the chosen model and compares the results with mean cost assignment. Section 3.6 offers concluding remarks.

3.2 Empirical methodology

3.2.1 Random forests

Generally, any supervised machine learning algorithm desires to predict a value (y) from its observed characteristics (x) via searching for a function $\hat{f}$ that minimizes a specified loss function $L(\hat{f}(x), y)$ on a new data point from the same distribution (Mullainathan & Spiess, 2017). For this chapter's case of the wildlife strike cost problem, x is various characteristics of the strike, including the type of plane, species struck, and phase of flight in which the strike occurred, while y is the direct repair costs. Specifically, random forests that predict a continuous variable begin with regression trees that seek to map characteristics x to predicted values $\hat{y}$ via a tree that splits at various nodes. At each node, the value of an individual characteristic determines a split, ultimately arriving at a terminal node or leaf where a prediction is made by averaging all data points that reach a given leaf.

For each available feature in a given dataset (e.g., whether structural damage to the aircraft was indicated), the algorithm splits the data into two groups, comparing the observed values of the target variable (e.g., direct repair costs due to wildlife strikes) to the corresponding sample mean of each subset using mean squared error (MSE). The feature with the lowest MSE is chosen for the initial split or the root of the tree, subsequently growing two branches that lead to nodes representative of the sample average for each respective group. The same iterative process of searching the remaining features is repeated for each new node in the tree. When a new split results in a lower MSE than the original split leading to a given node, the samples at that node are also split, growing branches to new nodes where the feature selection process is repeated. This recursive process continues until the stopping criteria are reached to avoid overfitting, as a tree with enough branches would predict each within-sample observation perfectly. Terminal nodes or leaves are reached that contain the average for all groups determined by the path of each branch, which begins at the root node and ends at the corresponding leaf of the tree. Thus, regression trees are theoretically better suited for data that is highly non-linear due to the foregoing process of partitioning the prediction space into small subspaces, instead of estimating a linear model across the entire space (Theodoridis & Koutroumbas, 2009).

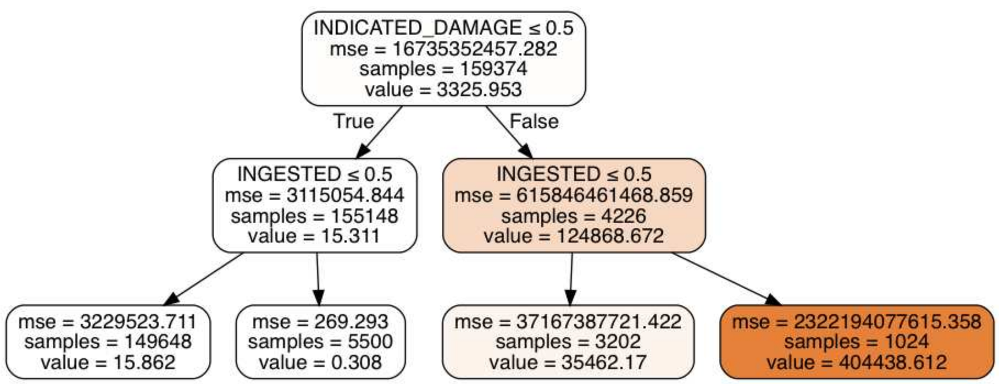


Figure 3.1: 2 feature regression tree: Example

Figure 3.1 visualizes the foregoing process using data from this chapter by fitting a simple regression tree on direct repair costs after a damaging wildlife strike using $n = 2$ binary features, namely whether structural damage to the aircraft was reported (INDICATED_DAMAGE) and whether the animal was ingested into the aircraft's engine (INGESTED). The algorithm selects INDICATED_DAMAGE as the tree's root, splitting the data into two unique groups or nodes conditional on the value of INDICATED_DAMAGE (0=No, 1=Yes). INDICATED_DAMAGE is chosen as the tree's root because when the data is split on INDICATED_DAMAGE, comparing each group's sample mean to their observed costs delivers a lower MSE on the whole than when initially splitting on INGESTED (0=No, 1=Yes). Because only 2 features are passed to the algorithm, it then checks at each new node if splitting on the remaining feature INGESTED provides an MSE lower when using the sample mean of the current node. As evident in the figure, this is the case for each new node as they are subsequently split on INGESTED, dividing the data further into four unique groups. These groups are the leaves or terminal nodes of the regression tree. Each leaf represents a unique subset of data; for example, the leftward-most leaf represents all events in the dataset that did not indicate damage and had no ingestion of an animal into the aircraft's engine. If this regression tree were to receive an out-of-sample observation that had the aforementioned characteristics, it would predict the corresponding leaf's mean of $15.86.

Random forests are an average over many regression trees to combat the previously stated empirical problem of overfitting associated with tree-based learning. According to Mullainathan and Spiess (2017), "[e]ach tree is fitted on a bootstrap sample of the ... set and constrained to a random subsample of features. The predictions of the trees are then averaged" (p.94). Specifically, a forest of regression trees is constructed such that each tree is fit on a unique set of features from all that are available. The leaves of each tree contain values representative of group averages determined by the recursive process of searching for splits across available features to fit the tree. Because each tree contains a unique subset of randomly selected features, the predictions delivered by a random forest for a single observation are the average across all relevant leaves of every tree in the forest and mitigate the previously mentioned overfitting problem linked to tree-based methods.

In addition to the nature of algorithms, random forests and other machine learning algorithms allow for regularization or hyperparameter tuning to combat overfitting and improve predictive performance. Such hyperparameters that can potentially be tuned in a random forest include but are not limited to the number of trees in the forest, number of variables used in each tree, depth of each tree, minimum number of samples required to split at a given node, and minimal leaf size (i.e., the number of observations required to allow a node to be terminal).

The empirical process of algorithm tuning using hyperparameters often incorporates what is broadly known as cross-validation. In its simplest form, cross-validation involves withholding a specified amount of data from the algorithm fitting process and adjusting hyperparameters based on the algorithm's performance on the withheld portion of the sample. However, this naive approach may be problematic in the sense that any set of hyperparameters selected for a given model may be dependent on the portion of the withheld data, limiting model generalization to actual out-of-sample data. One way to address this issue is implementing k-fold cross-validation.

Data available for fitting[25] is first randomly partitioned into k unique sections or folds. Next, for all folds f in $1, \ldots, k$ the algorithm is fit to $k-1$ folds $\neq f$ and the out-of-sample accuracy evaluated on fold f. This can be done with various combinations of hyperparameters, ultimately selecting the set with the best average performance across all folds. This chapter uses a modified version of k-fold that incorporates a validation set in addition to each withheld fold (or test set) for both the random forest and baseline regression model. It also incorporates recursive feature elimination (RFE) to assist with variable selection. Assuming a researcher has n potential features available for use in an algorithm, RFE fits a model with the 'best' n features for all n as ranked, among others, by coefficient size and the mean impurity decrease. A specified loss is then calculated for each of the n models, and the set of features with the lowest loss is selected as the optimal variable subset.

Combining these two components, the following tuning process was used to generate the random forest results presented in Section 3.4:

1. All data available to fit the algorithm is split into two unique portions, 'train' and 'other.'

[25]Where the true value of the target variable (y) is observed.

2. The ’other’ portion of data is then split into validation and test sets.
3. Models fit using the training data that employ different combinations of hyperparameters are grid searched and selected based on out-of-sample accuracy on the validation set. Mean squared error is chosen to measure all accuracy in the analysis. This chapter searches 81 different combinations of models by varying three choices for the number of trees in the forest, the depth of each tree (how many nodes in a single tree), leaf size, and the minimum number of samples required to split at a given node. Only these 4 are searched due to computational limitations.
4. RFE is performed on the selected hyperparameter-tuned model from the previous step, similarly choosing the best subset of features based on validation set accuracy.
5. The final model containing both the selected hyperparameters and features is evaluated based on test set accuracy.
6. Test set accuracy is recorded, and subsequently, the validation and test sets are switched.
7. Steps 3-5 are repeated using the new validation and test sets.
8. Steps 1-7 are repeated for a new, unique combination of ’train’ and ’other’ sets.
9. Steps 1-8 are repeated a total of 3 times, resulting in 6 unique model accuracy scores. Final model performance is ultimately judged based on the average of the 6 scores (or ’iterations’).

3.2.2 Neural networks

An artificial neural network or, simply, ’neural network,’ is a supervised machine learning algorithm, composed of several layers organized hierarchically. The first layer is the input layer, which receives a vector of descriptive features x. Again, in this chapter these would be characteristics of a damaging wildlife strike event, such as whether the animal was ingested into the aircraft’s engine. Subsequent intermediate layers, known as hidden layers, ultimately transform the inputs into a prediction, which is, for this chapter, direct repair costs from wildlife strikes ($\hat{y}$) in the final or output layer. Each layer is composed of several different neurons or hidden units with designated activation functions. Activation functions compress weighted-linear combinations of inputs ($wx + b$, where b is called the bias term) as information flows through each layer. The inputs for each layer are the resulting values from this compression process of the previous layer. Figure 3.2 visualizes a *2* hidden layer neural network taking six input characteristics, with neurons represented by the white circles in each respective layer.

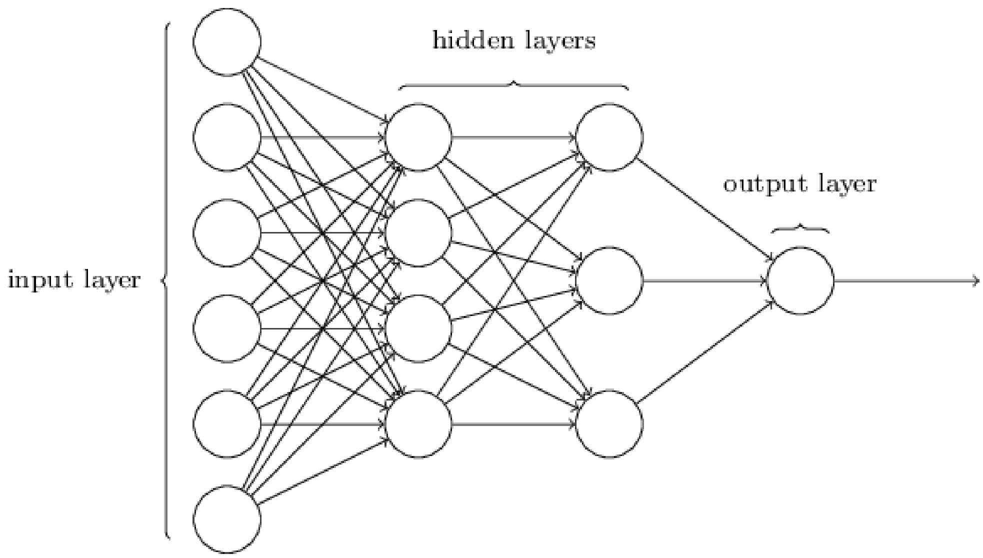


Figure 3.2: 2-hidden-layer neural network: Example

The aforementioned process is repeated iteratively in conjunction with gradient descent to find a vector of weights (w) that minimize the loss function $L(...)$, which is based on comparing predictions from the network ($\hat{y}$) with actual values (y). Specifically, the errors from the loss function are fed backward through the network to find optimal weights, updating each value by an amount proportional to the derivative of the loss function with respect to a given weight.

For a formal presentation of the example in Figure 3.2 (assuming a sample size of m), let

- n_x represent the number of input features, in this case $n_x = 6$,

- L be the *total* number of layers (indexed by superscript [l]) in the network: $l = 0, ..., L$ (for this example, $L = 3$),
- Superscript (i) denote the i^{th} observation for sample size m: $i = 1, ..., m$,
- Subscript h index the hidden unit number in layer [l], for $h = 1, ..., n^{[l]}$ (e.g., $a_1^{[1]}$ represents an activation value associated with the first hidden unit [neuron] in the first layer with $n^{[l]} = 4$ total neurons), and
- $g_h^{[l]}(...)$ the specified activation function for hidden unit h in layer l.

First, a matrix X of dimensions (n_x, m) where input features are indexed vertically and observations indexed horizontally is passed to the first hidden layer ($[l] = 1$) of the network. Each hidden unit in this layer receives the matrix X and creates a weighted linear combination of these inputs for $h = 1, ..., 4$:

$$Z_h^{[1]} = w_h^{[1]} X + b_h^{[1]}$$

Each $Z_h^{[1]}$ is transformed non-linearly through a chosen activation function to deliver a $(n^{[1]}, m)$ matrix $A^{[1]}$ of values to be passed on to each hidden unit in the following layer. In this chapter, all activation functions are a rectified linear unit (ReLU), except in the output layer, which is linear due to the nature of the prediction. ReLu is one of the most commonly used activation functions in deep learning, transforming all elements x of $Z_h^{[l]}$ to positive values such that $f(x) = \max(0, x)$ (Ramachandran, Zoph, & Le, 2017). The next layer ($[l] = 2$) performs a similar process, creating corresponding $Z_h^{[2]}$ for $h = 1, ..., 3$ values to be passed though a specified activation function, where:

$$Z_h^{[2]} = w_h^{[2]} A^{[1]} + b_h^{[2]}$$

The resulting set of values $A^{[2]}$ is an ($n^{[2]}$, m) matrix passed to the final output layer, delivering a (1, m) dimensional vector of predictions for each observation ($\hat{y}$) through an activation function ($g^{[3]}(Z^{[3]})$) of the values[26]:

$$Z^{[3]} = w^{[3]}A^{[2]} + b^{[3]}$$

This process, known as forward-propagation, can be expressed generally by the following two equations for a neural network with hidden layers $l = 1, ..., L$:

$$Z^{[l]} = w^{[l]}A^{[l-1]} + b^{[l]} \tag{3.1a}$$

$$A^{[l]} = g^{[l]}(Z^{[l]}) \tag{3.1b}$$

Where $Z^{[l]}$ and $A^{[l]}$ are ($n^{[l]}, m$) vectors, while the weight matrix $w^{[l]}$ is of dimensions ($n^{[l]}, n^{[l-1]}$), and the bias term $b^{[l]}$ is a ($n^{[l]}, 1$) matrix.

The prediction vector $\hat{y}$ is combined with the observed y values to calculate the chosen loss function, $L(...)$. Following this calculation, and using the process of gradient descent, partial derivatives of the loss function with respect to the weight and bias matrices of each hidden layer are calculated, and parameters are updated from their initial values according to a specified learning rule. Because one pass forward via forward propagation and backward using gradient descent through the network is a single iteration, this process is repeated iteratively to find the entire set of parameters w and b for all layers $l = 1, ..., L$ that minimize the loss function.

Similar to a random forest, several hyperparameters can be tuned to improve the prediction performance of a neural net. These hyperparameters include the number of hidden layers, number of specific neurons in each hidden layer (which need not be the same), proportional updating rule with respect to the partial derivative of the cost function (i.e., the learning rate), activation function for specific neurons, and loss function specification, as well as the number of iterations to go forward and backward through the network.

[26]Note that the h notation is dropped; there is only a single hidden unit in the output layer.

Due to computational constraints, the neural network estimated in Section 3.4 is fit via a less mechanized process than the random forest or baseline regression model:

1. A set of hyperparameters and network structure are specified.
2. All available data for fitting is split into train and test sets.
3. The model is fit using the training set, while accuracy is subsequently evaluated based on test set performance.
4. Steps 1-3 are repeated a total of 6 times, resulting in 6 unique model accuracy scores. Final model performance is again judged based on average accuracy across the 6 scores (or iterations).
5. Hyperparameters and network architecture are tweaked as needed to improve final model performance.
6. Steps 1-5 are repeated until a final hyperparameter set and network architecture are selected.

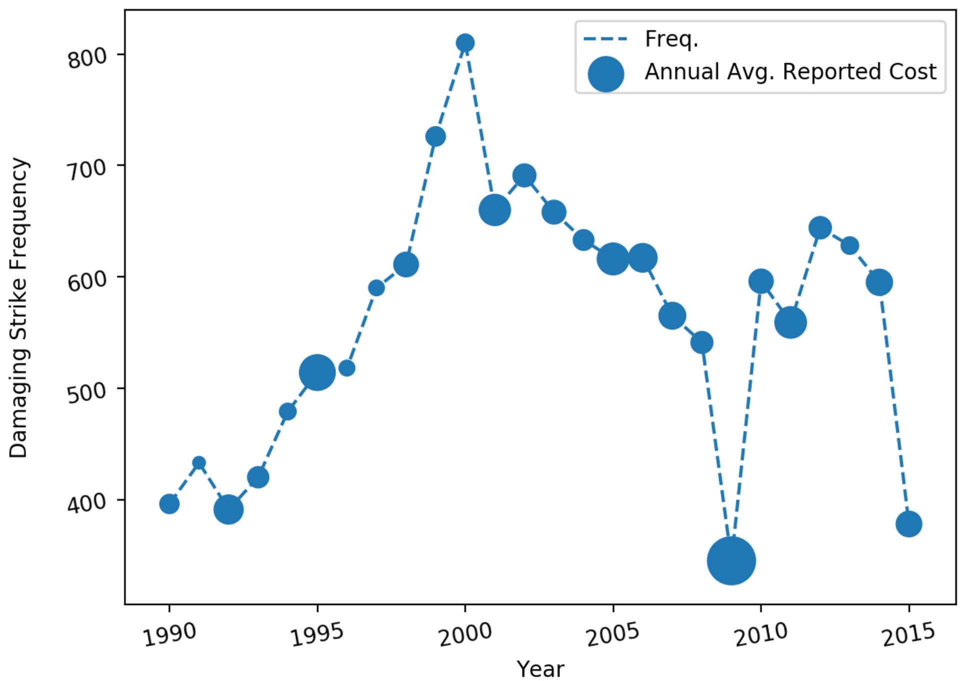
The final neural network used consists of two hidden layers in addition to its input (149 [27] neurons) and output (1 neuron) layers with 75 and 25 neurons, respectively. All activation functions between input and hidden layers are a rectified linear unit (ReLU). The activation function in the output layer is linear due to the continuous nature of our predicted variable.

3.3 Data

The NWSD currently contains records of all wildlife strikes in the US to civilian and military airplanes reported on a voluntary basis from 1990 to June 2018 at non-military airports (*Federal Aviation Administration National Wildlife Strike Database*, 2018). Figure 3.3 provides an overview of the frequency of damaging wildlife strikes in the database

[27] This number is greater than the number of *unique* variables listed and discussed in the following section because all categorical inputs to a neural network must be binary; this is reflected in Table D.1 of the Appendix.

(i.e., those that reported structural damage to the aircraft), as well as the average reported repair cost associated with such strikes over the 1990-2015 time period.[28] While the NWSD contains some strikes from 2016-2018, there was insufficient cost data reported to calculate a meaningful annual average and are thus omitted from the figure. The size of the bubble is representative of the continuous value of the annual average damage cost in USD. Damaging strike frequency appears to have peaked at the beginning of the new millennium, followed by a decline. It should be noted that this peak may be reflective of real increases in reporting behaviors.



Source: NWSD

Figure 3.3: Damaging wildlife-strike frequency and annual average reported repair costs: 1990-2015

[28]Dolbeer (2015) estimates damaging strike reports in the NWSD account for approximately 78% and 91% of actual occurrences for the 2004-2008 and 2009-2013 time periods, respectively.

At the time of access, there were a total of 169,726 total observations in the database. This chapter assumes no reported structural damage to the aircraft implies a zero damage cost strike. There are 159,374 observations available to fit or train our models (i.e., the training set), meaning we possess necessary information related to our target variable of interest: direct repair costs.

Table 3.1: Training set summary statistics: reported repair costs (USD)

	Mean	*Std.*	*Min.*	*Max.*
COST_REPAIRS	3325.95	129365.60	0.0	36000000.00

Table 3.1 presents summary statistics related to this direct cost variable. The average reported cost in our training set, including costless strikes, is approximately $3,326, carrying a rather high standard deviation close to $129,366. The average non-zero damage cost in our training set is approximately $123,790, while the maximum reported repair cost is $36,000,000. As evident in Figure 3.4, the cost variable within our training data is heavily skewed right.

The full NWSD contains information on over 90 different variables associated with each individual strike. Features (i.e., x characteristics) used in our analysis were selected in an effort to (1) maximize the sample size and (2) encompass as many factors that may influence the cost of a wildlife strike as possible.[29] The selected features include the aircraft type (AC_CLASS), aircraft mass (AC_MASS), number of birds seen by the aircraft pilot (BIRDS_SEEN), number of birds struck (BIRDS_STRUCK), effect of the strike on the flight (EFFECT), position of the first (ENG_1_POS) through the fourth engine (ENG_4_POS), number of engines on the aircraft (NUM_ENGS), phase of flight when the strike occurred (PHASE_OF_FLT), precipitation (PRECIP), size of the animal struck (SIZE), cloud cover (SKY), time of day of the strike event (TIME_OF_DAY), engine

[29] Similar to Anderson et al. (2015).

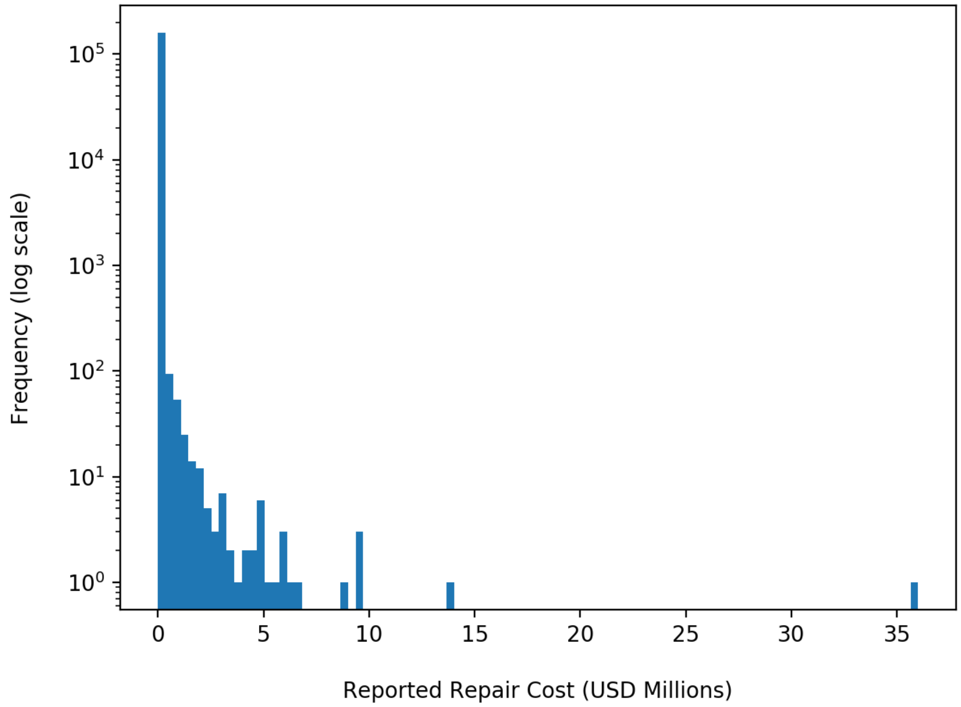


Figure 3.4: Training set distribution of reported repair costs

type of the aircraft (TYPE_ENG), and whether the pilot was warned of birds or wildlife prior to the strike (WARNED). Individual binary variables for the specific aircraft component struck (STR_) include the aircraft component damaged (DAM_), whether structural damage to the aircraft was reported (INDICATED_DAMAGE), and whether the animal was ingested by the aircraft's engine (INGESTED). A categorical damage category variable (DAMAGE_) is included. This amounts to a total of 48 unique variables available to train our models. Detailed summary statistics for all training set features are available in Table D.1 of the Appendix.

3.4 Model selection

Table 3.2 presents a comparison of the two machine learning algorithms, random forest and neural network, to the baseline regression model. Visualized in Figure 3.5, the neural network appears to be the best performing model of the three examined, improv-

Table 3.2: Different algorithms' performance in predicting wildlife strike damage costs

Model	*Average test root mean squared error*	*% Relative improvement over baseline*
OLS w/ RFE (Baseline)	66412.571431	-
Random Forest w/ RFE	68796.022941	-3.59%
Neural Network	62268.253573	6.24%

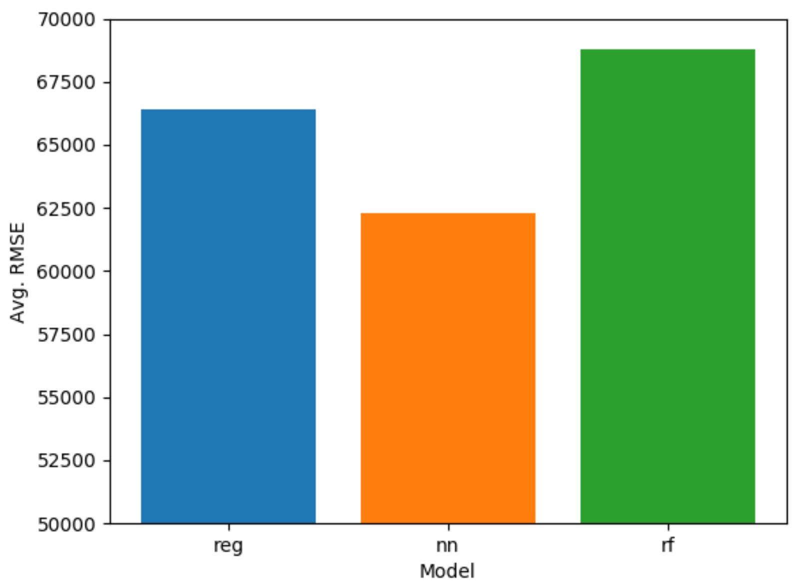


Figure 3.5: Average test set RMSE across different algorithms

ing accuracy from the baseline by approximately 6.25%. Random forest fails to beat the baseline model in terms of performance; it is approximately 4% less accurate than the regression model.

Figure 3.6 plots each unique iteration score for all models. It is obvious that the neural network consistently has the lowest root mean squared error, or the best performance across all 6 iterations. While the primary focus for judging algorithm performance in this

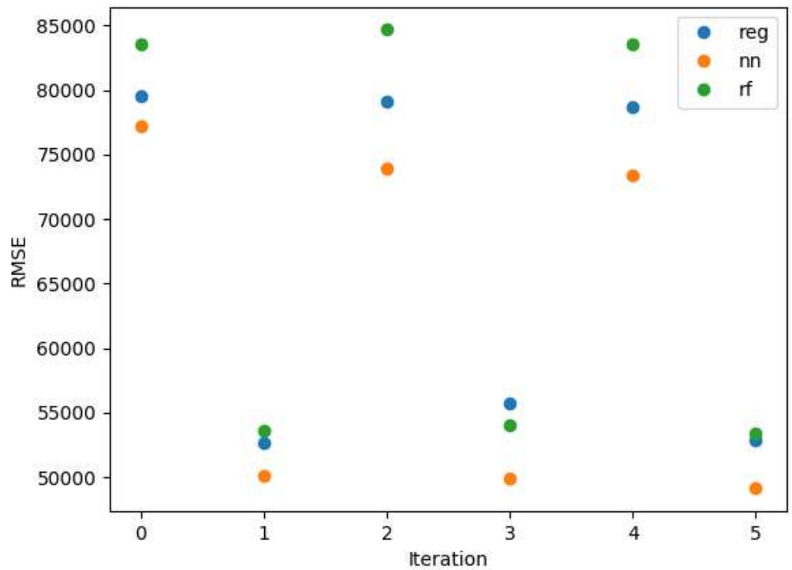


Figure 3.6: Test set RMSE across different algorithms by iteration

chapter is RMSE, it could the other algorithms could perform best when judged under different criteria. To informally test this hypothesis, test set R^2 and mean absolute error were also tracked across all iterations. Shown in Figure 3.7, and mirroring the results associated with RMSE, the neural network has the highest or best R^2 on average across all algorithms. Plotting iteration-specific results in Figure 3.8, the neural network has the highest R^2 across every instance.

The theme of relatively superior performance by the neural network continues when examining mean absolute error. Evident in Figure 3.9, the neural network also possesses the lowest MAE on average across all iterations. In Figure 3.10, which plots the results across iterations, the neural network posses the lowest MAE in 5 out of 6 total iterations.

Although variation in performance across models is the main focus of this section, of additional interest are the features selected for each iteration via RFE for the baseline regression and random forest algorithms. As described previously, features are selected for each iteration based on the validation set performance of the aforementioned models. Figure 3.11 shows a total count of features selected across all iterations for the baseline

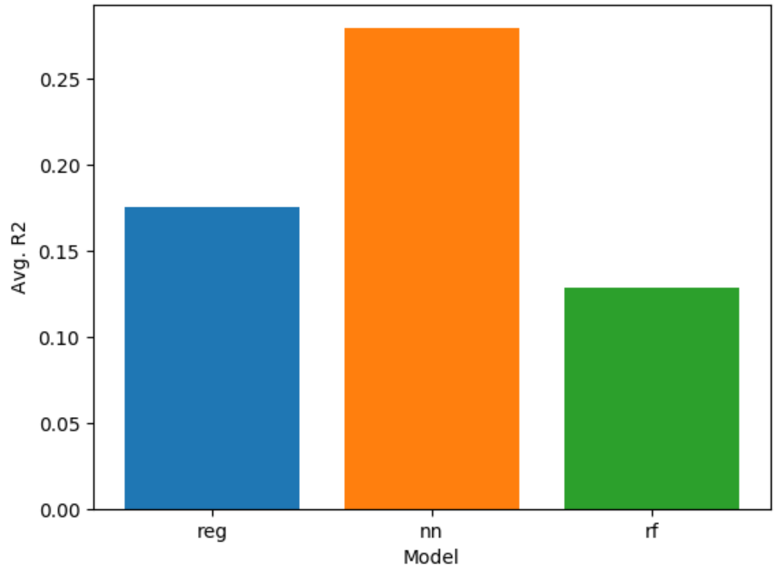


Figure 3.7: Average test set R^2 across different algorithms

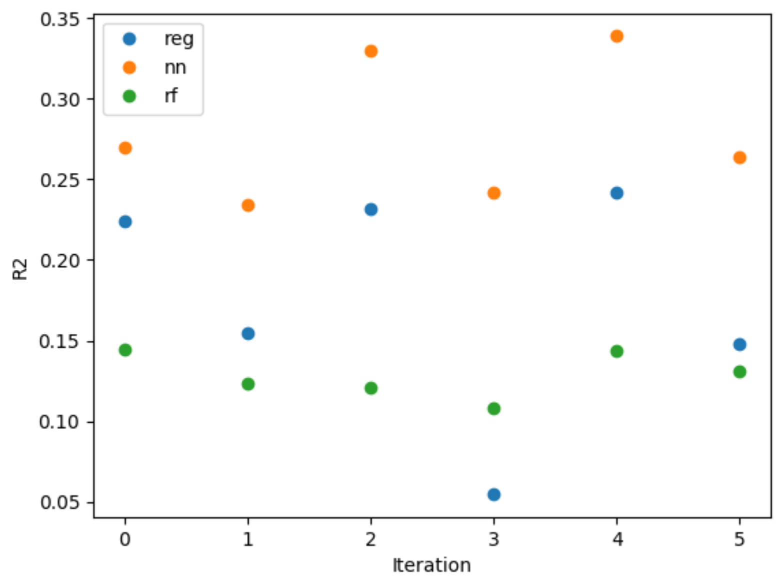


Figure 3.8: Test set R^2 across different algorithms by iteration

and random forest models where RFE was used. Unsurprisingly, the most important strike characteristics across the two models satisfy any naive *a-priori* expectations. Across both models and iterations, the maximum a feature could feasibly be selected is 12 times. Cat-

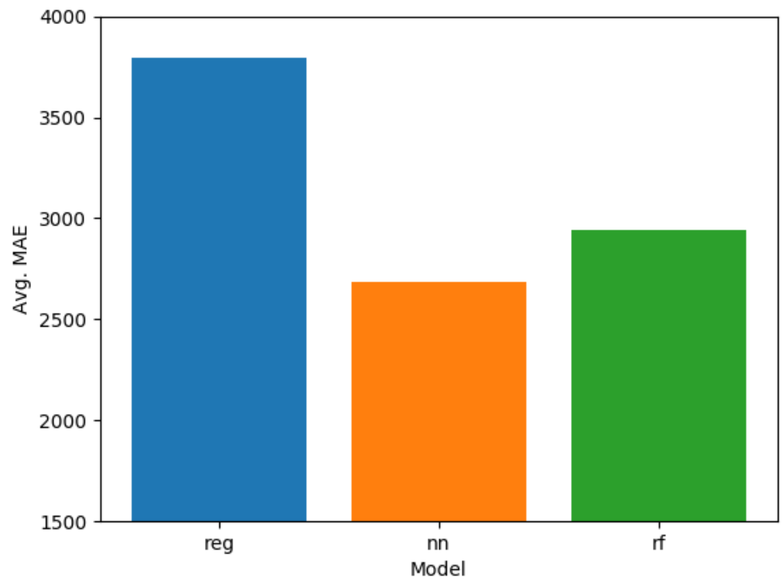


Figure 3.9: Average test set MAE across different algorithms

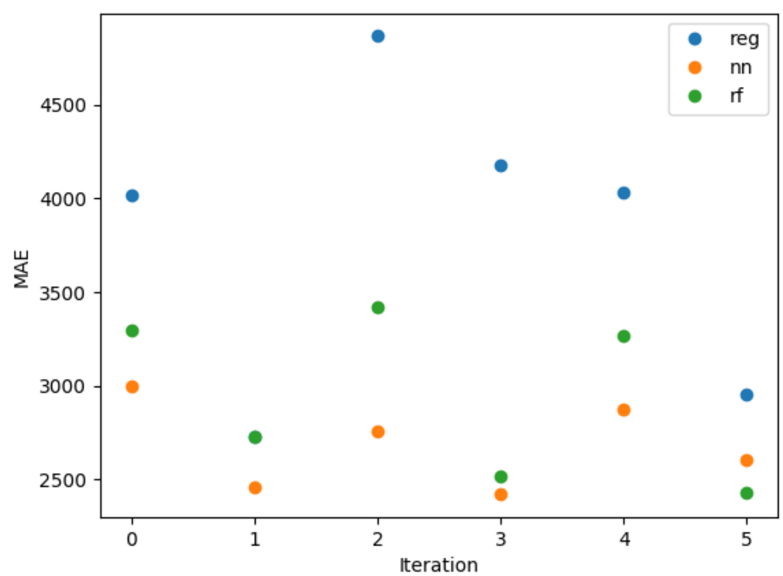


Figure 3.10: Test set MAE across different algorithms by iteration

egorical damage to the aircraft (DAMAGE_), damage to the engine (DAM_ENG), whether the strike caused engine failure (EFFECT_ENGINE_SHUTDOWN), and whether structural damage to the aircraft was indicated (INDICATED_DAMAGE) were all selected close to 10 times across both the baseline regression and random forest models.

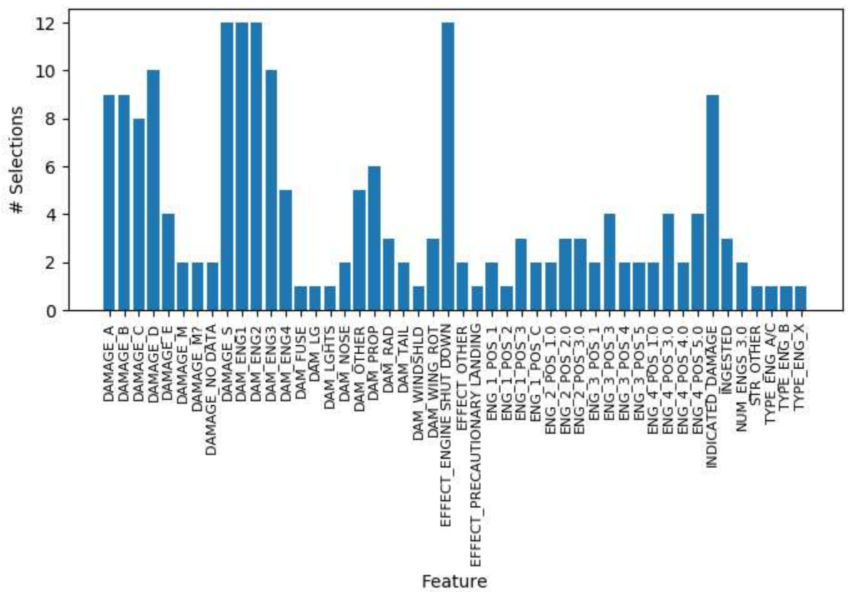


Figure 3.11: RFE-selected feature count: Baseline and random forest models

3.4.1 Sensitivity analysis

Considering the skewed distribution of the reported repair cost variable in this chapter, those previously presented in this section could be contingent on the presence of outliers. Specifically, it could be the case that the neural network outperforms the baseline regression and random forest models in the presence of outliers. Once the outliers are removed, the results presented could plausibly change. To test this proposition and provide a sensitivity analysis, the process conducted previously is repeated while removing all aircraft that are labeled 'destroyed' in the original training set. More specifically, we removed observations in which DAMAGE_D equals 1.

Table 3.3 presents average test set RMSEs across all iterations for the three models following the removal of destroyed aircraft. The results generally match those presented in the prior section. The neural network appears to perform the best, while the random forest

Table 3.3: Different algorithms' performance in predicting wildlife-strike damage costs: no destroyed aircraft

Model	*Average test root mean squared error*	*% Relative improvement over baseline*
OLS w/ RFE (Baseline)	65946.0974821	-
Random Forest w/ RFE	69594.013271	-5.53%
Neural Network	62001.424886	5.98%

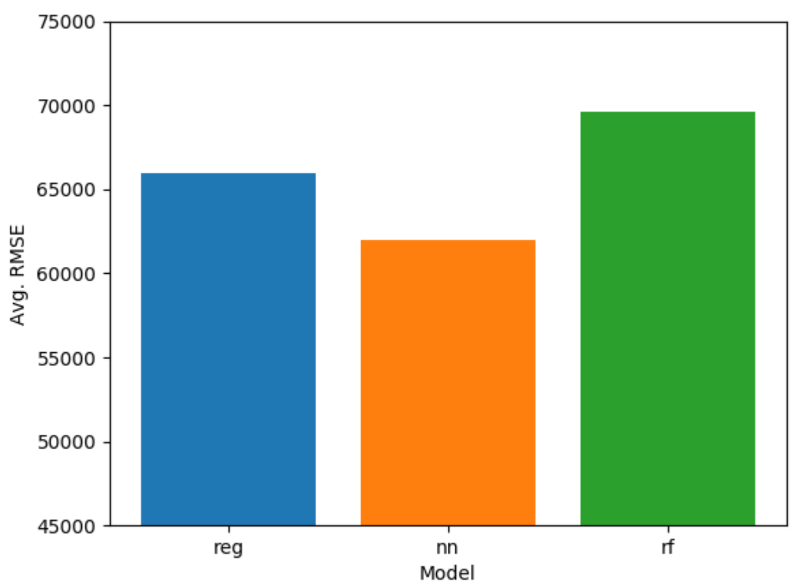


Figure 3.12: Average test set RMSE across different algorithms: No destroyed aircraft

again fails to outperform the baseline model. These findings are visualized in Figure 3.12. Figure 3.13 expands the RMSE results across iterations, indicating that the neural net consistently possesses the best test set performance. A similar result is shown for R^2 in Figures 3.14 and 3.15, as well as for MAE, shown in Figures 3.16 and 3.17. MAE, specifically, is improved for the neural net relative to the inclusion of outliers, which now possesses the lowest MAE in 6 out of 6 iterations. This sensitivity analysis reinforces

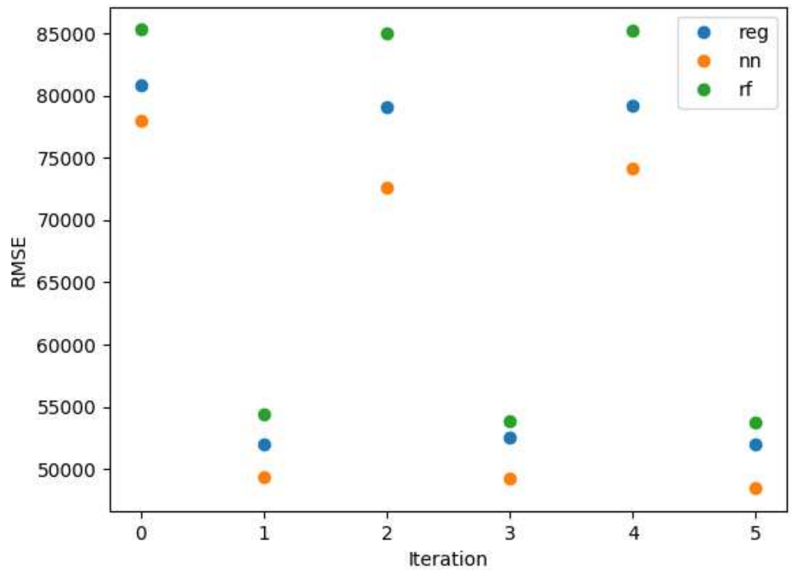


Figure 3.13: Test set RMSE across different algorithms by iteration: No destroyed aircraft

the findings of the primary exercise: the neural net is the best performing algorithm for missing cost prediction in the NWSD of the three tested models, both in the presence and absence of outliers.

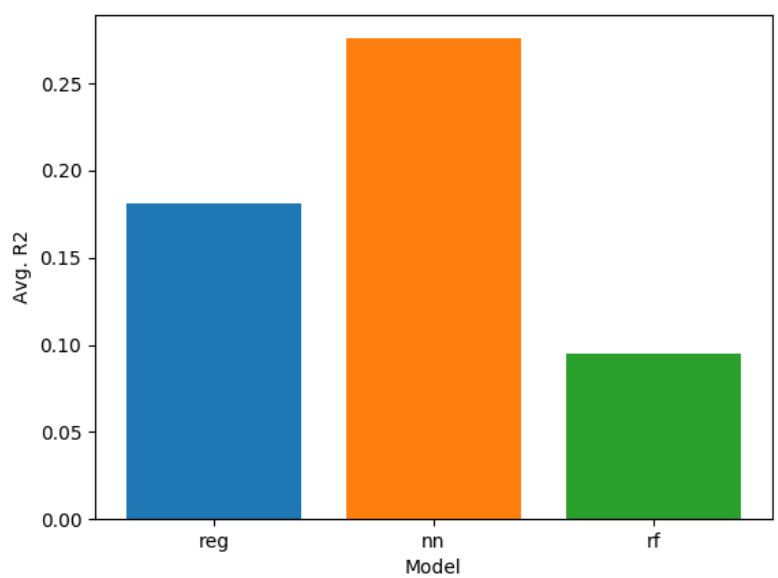


Figure 3.14: Average test set R^2 across different algorithms: No destroyed aircraft

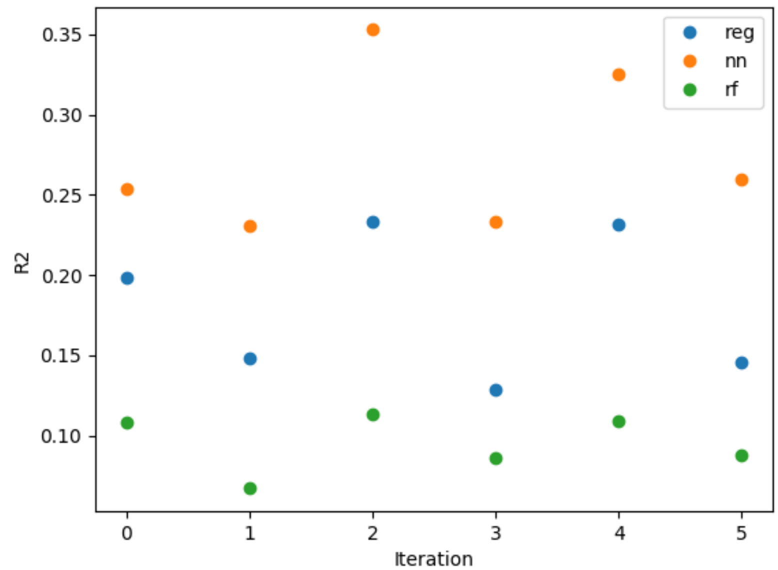


Figure 3.15: Test set R^2 across different algorithms by iteration: No destroyed aircraft

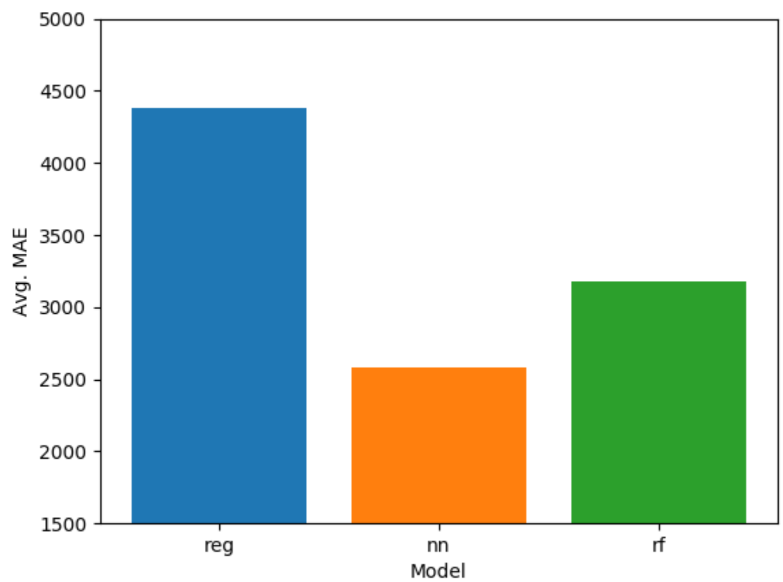


Figure 3.16: Average test set MAE across different algorithms: No destroyed aircraft

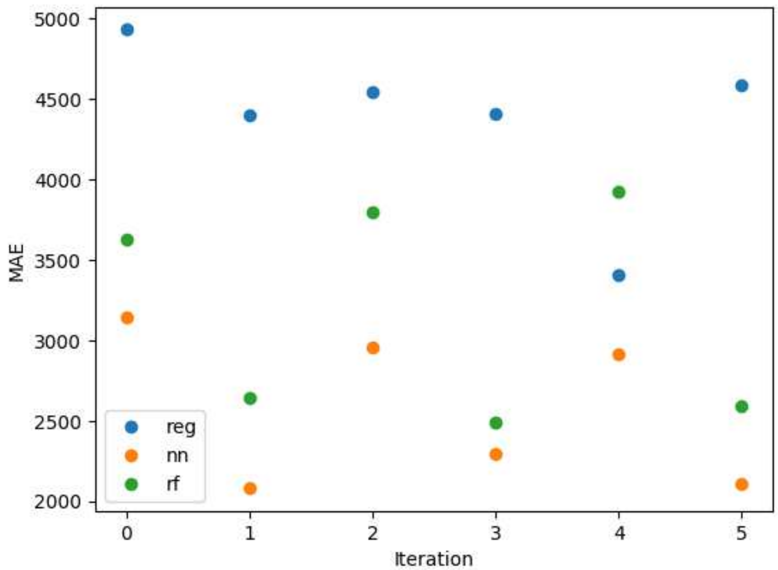


Figure 3.17: Test set MAE across different algorithms by iteration: No destroyed aircraft

3.5 Application

The previous section shows the neural network is the best of the three models at predicting missing costs in the NWSD. This section first applies this model to 10,388 observations from the NWSD that indicate structural damage to the aircraft but fail to report an associated repair cost.[30] Next, aggregate measures are derived that correspond to the direct economic costs of wildlife strikes in the United States. These estimates are then compared to estimates obtained by using mean cost assignment. Costs are interpolated using both methods only for the 1990-2014 time period to make comparisons to Dolbeer et al. (2014)'s $157 million average annual estimate more relevant.

Dolbeer et al. (2014)'s measure is derived by assuming all strike events indicating an adverse effect on a flight were damaging, then assigning the average reported cost in a given year to all strikes labeled as such. The mean cost assignment approach taken in this section is slightly different in that only strikes that explicitly indicate structural damage

[30]See Section 3.3.

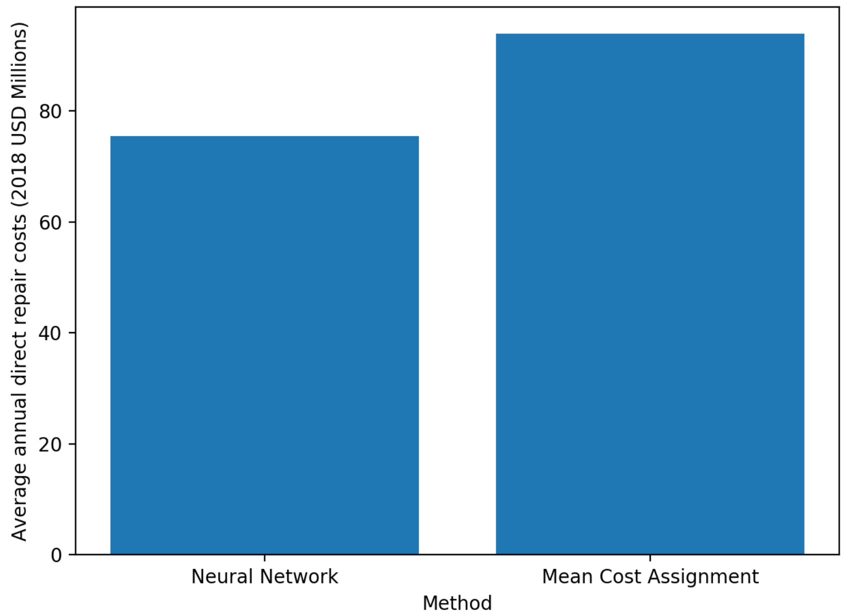


Figure 3.18: Estimated average annual direct costs of wildlife strikes in the US, 1990-2014

to the aircraft are examined. This is done to avoid biasing the estimates upward as an adverse effect on the flight does not directly imply costly aircraft damage. It should also be noted that Dolbeer et al. (2014) state they only use 'inflation-adjusted' costs but do not provide any context for replication. For this reason, a novel version of mean cost assignment is devised to allow for meaningful comparison between methods.

For each year in the 1990-2014 range, the missing costs are given either the average reported direct repair cost or the neural net's prediction in the corresponding year. All costs for each year are summed, and the annual total is adjusted to represent 2018 US dollars (USD). Figure 3.18 shows the main results from the foregoing process for both the neural network and mean cost assignment methods. The aggregate average annual repair cost burden across 1990-2014 is approximately $19 million, or 25% lower for the neural network than mean cost assignment. Specifically, the neural network estimates imply the average national direct repair costs to be $74,705,137.62, while mean cost assignment estimates imply the same measure to be $93,944,431.08. This large discrepancy between the two reported estimates highlights the problematic nature of using mean

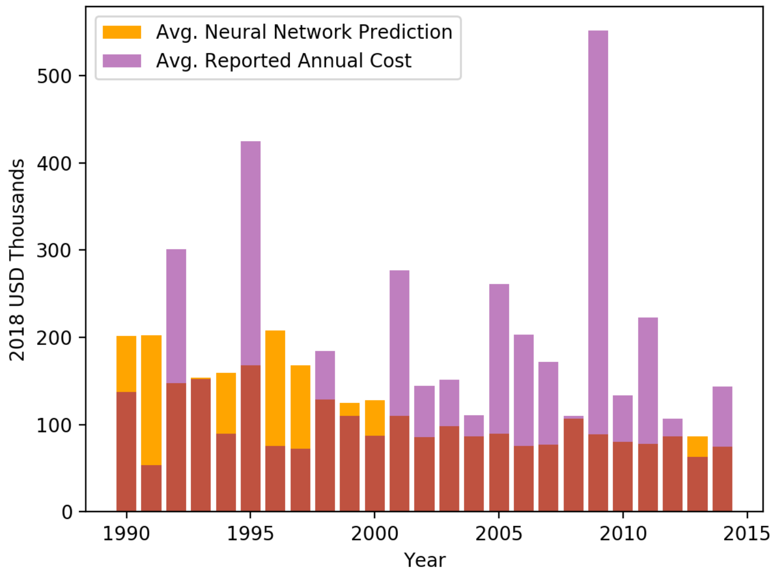


Figure 3.19: Average neural network cost prediction versus average reported cost, 1990-2014

cost assignment. The approximately $82 million difference between the neural network's estimate and Dolbeer et al. (2014)'s $157 million reinforces the issues with previous naive methods.

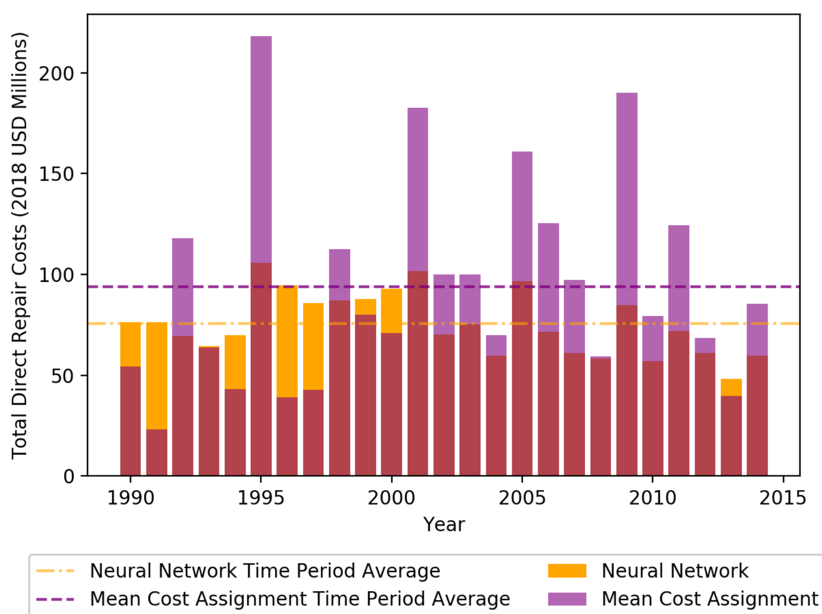


Figure 3.20: Neural network versus mean cost assignment: Total direct repair costs, 1990-2014

Figure 3.19 plots the average reported cost in thousands of dollars and the average predicted cost by the neural network for each year. These costs are then adjusted to represent 2018 USD. While no obvious temporal trend is evident, across all years, the average reported repair cost is higher than the mean neural net prediction for most years. Figure 3.20 decomposes the results in Figure 3.18 across the years. As expected, the annual total direct repair cost burden is higher when mean cost assignment is used than the neural network predicts across most of the years.

The primary takeaway from this section builds on the results presented previously: while the neural network emerged as the best predictive algorithm, this application highlights the differences and dangers of naive empirical methods used to interpolate missing information in the NWSD and, potentially, elsewhere. Both the $19 and $82 million dollar differences between the neural network and mean cost assignment's estimates are presumably a resource that can be allocated elsewhere more efficiently as wildlife management resources are finite. Additionally, this model can be used effectively at the airport level to predict damage costs representative of damaging strikes a given airport observes in reality.

3.6 Conclusion

The first results in this chapter show machine learning algorithms, specifically a neural network to provide a non-trivial improvement in performance relative to simple ordinary least squares. Additionally, this conclusion appears robust to the presence of outliers. There are many future steps to be taken in this realm, most importantly, searching over a larger set and variety of hyperparameters to ensure the best possible predictive models. Another is to incorporate additional machine learning algorithms omitted from this analysis, including k-nearest neighbors, least absolute shrinkage and selection operator (LASSO), and gradient boosted regression, among others.

The second set of results presented highlight the importance of model development to more accurately predict missing costs in the NWSD. Interpolating missing costs with a neural net delivers an average annual direct repair cost estimate from wildlife strikes in the US that is $19 million dollars lower than when using mean cost assignment. Moreover, it is $82 million dollars lower than Dolbeer et al. (2014)'s reported estimate. This information is relevant to policymakers because they could use it to adjust funding and more efficiently improve performance in the face of limited resources. Additionally, the development of such a model allows for more accurate program evaluation at the airport level, allowing airports to estimate damage based on what is observed in reality instead of an airport or national average.

The goal of the ongoing research presented in this chapter continues the development of the most accurate predictive tool to ultimately assist in guiding the efficient allocation of finite management resources in the face of wildlife-aircraft conflict.

Conclusion

This dissertation has broadly investigated direct repair costs associated with wildlife strikes to aircraft in the United States. Chapter 1 and Chapter 2 examine the likelihood and accuracy of cost disclosure by large domestic American airlines, who voluntarily provide such information to the NWSD. Results show that these behaviors are liked to market competition and profitability. For a given airline, both the probability of disclosure and accuracy when a cost is disclosed are decreasing with respect to market competition and firm profitability. This information is beneficial to any future decision-maker associated with policy linked to wildlife strike cost disclosure.

Chaper 3 focuses on method development for interpolation of missing repair costs in the NWSD. Results show a neural network to outperform traditional regression and a random forest model when predicting out-of-sample data. Moreover, the issues with previous naive methods used to develop national estimates of the economic burden caused by wildlife-aircraft conflict are explained. Specifically, predicting missing costs using a neural network delivers an average annual estimate of the direct costs of wildlife strikes in the US that is approximately \$75 million–\$19 million lower than when using mean cost assignment and \$82 million lower than Dolbeer et al. (2014)'s estimate using a variation of the same method. In the face of this discrepancy, while highlighting the dangers of using naive empirical methods, policymakers could presumably allocate funding more efficiently. This research also provides a potential tool for program efficiency evaluation at the airport level by allowing airports to estimate damage based on what is observed in reality instead of an airport or national average.

References

Airlines for America. (2018). *Airlines for America glossary of airline terminology.* Online. Retrieved from http://airlines.org/glossary/

Airlines for America. (2016). *Status of air travel in the USA.* Online. Retrieved from http://airlines.org/wp-content/uploads/2016/04/2016Survey.pdf

Allan, J. R. (2000). The costs of bird strikes and bird strike prevention. In *Human conflicts with wildlife: Economic considerations.* USDA National Wildlife Research Center Symposia. Online. Retrieved from http://digitalcommons.unl.edu/nwrchumanconflicts/18/.

Anderson, A., Carpenter, D. S., Begier, M. J., Blackwell, B. F., DeVault, T. L., & Shwiff, S. (2015). Modeling the cost of bird strikes to US civil aircraft. *Transportation Research Part D*, *38*, 49-58.

Armantier, O., & Richard, O. (2003). Entry and exchanges of cost information. *Journal of Regulatory Economics*, *24*(2), 223-241.

Arya, A., & Mittendorf, B. (2007). The interaction among disclosure, competition between firms, and analyst following. *Journal of Accounting and Economics*, *43*, 321-339.

Baum, C. F. (2013). *Implementing new econometric tools in Stata.*

Begier, M. J., & Dolbeer, R. A. (2009). *Protecting the flying public and minimizing economic losses within the aviation industry* (Tech. Rep.). Washington DC: United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services.

Benford, F. (1938). The law of anomalous numbers. In *Proceedings of the American Philosophical Society* (*78*), 551-572.

Board, O. (2009). Competition and disclosure. *Journal of Industrial Economics*, *57*(1), 197-213.

Botosan, C. A., & Stanford, M. (2005). Managers' motives to withhold segment disclosures and the effect of SFAS no. 131 on analysts' information environment. *The Accounting Review*, *80*(3), 751-777.

Brander, J. A., & Zhang, A. (1990). Market conduct in the airline industry: An empirical investigation. *The RAND Journal of Economics*, *21*(4), 567-583.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics*. New York: Cambridge.

Cho, W. K. T., & Gaines, B. J. (2007). Breaking the (Benford) law. *The American Statistician*, *61*(3), 218-223.

Clinch, G., & Verrecchia, R. E. (1997). Competitive disadvantage and discretionary disclosure in industries. *Australian Journal of Management*, *22*(2), 125-137.

Conroy, T., Low, S. A., & Weiler, S. (2017). Fueling job engines: Impacts of small business loans on establishment births in metropolitan and nonmetro counties. *Contemporary Economic Policy*, *35*(3), 578-595.

Cooper, J. C. B. (1999). Artificial neural networks versus multivariate statistics: An application from economics. *Journal of Applied Statistics*, *26*(7), 909-921.

Darrough, M. N. (1993). Disclosure policy and competition: Cournot vs. Bertrand. *The Accounting Review*, *68*(3), 534-561.

Darrough, M. N., & Stoughton, N. M. (1990). Financial disclosure policy in an entry game. *Journal of Accounting and Economics*, *12*, 219-243.

Davies, S. (1979). Choosing between concentration indices: The iso-concentration curve. *Economica, New Series*, *47*(181), 67-75.

Dedman, E., & Lennox, C. (2009). Perceived competition, profitability, and the withholding of information about sales and cost of sales. *Journal of Accounting and Economics*, *48*, 210-230.

de Marchi, S., & Hamilton, J. T. (2006). Assessing the accuracy of self-reported data: An evaluation of the toxics release inventory. *Journal of Risk and Uncertainty*, *32*, 57-76.

Desgranges, G., & Gauthier, S. (2016). Rationalizability and efficiency in an asymmetric

Cournot oligopoly. *Journal of Industrial Organization*, *44*, 163-176.

DeVault, T. L., Belant, J. L., Blackwell, B. F., & Seamens, T. W. (2011). Interspecific variation in wildlife hazards to aircraft: Implications for airport wildlife management. *Wildlife Society Bulletin*, *35*(4), 394-402.

DeVault, T. L., Blackwell, B. F., & Belant, J. L. (Eds.). (2013). *Wildlife in airport environments: Preventing animal-aircraft collisions through science-based management*. Baltimore: The John Hopkins University Press.

Dolbeer, R. A. (1998). Evaluation of shooting and falconry to reduce bird strikes with aircraft at John F. Kennedy international airport. In *International bird strike committee ibsc24/wp13.* Online. Retrieved from https://nwrc.contentdm.oclc.org/digital/collection/p16473coll8/id/16027/.

Dolbeer, R. A. (2015). *Trends in reporting of wildlife strikes with civil aircraft and in identification of species struck under a primarily voluntary reporting system, 1990-2013* (Tech. Rep.). Washington DC: United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services.

Dolbeer, R. A., & Wright, S. E. (2008). Wildlife strikes to civil aircraft in the united states 1990-2007. In *Other bird strike and aviation materials.* Lincoln: Bird Strike Committee Proceedings.

Dolbeer, R. A., & Wright, S. E. (2009). Safety management systems: how useful will the FAA National Wildlife Strike Database be? *Human-Wildlife Conflicts*, *3*(2), 167-178.

Dolbeer, R. A., Wright, S. E., Weller, J. R., Anderson, A. L., & Begier, M. J. (2014). *Wildlife strikes to civil aircraft in the United States, 1990-2014* (Tech. Rep.). United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services. Online. Retrieved from http://i2.cdn.turner.com/cnn/2016/images/09/07/faa.bird.strike.report.2014.pdf.

Dumas, C. F., & Devine, J. H. (2000). Detecting evidence of non-compliance in self-reported pollution emissions data: An application of Benford's law. In *2000 Annual* *meeting, July 30-August 2, Tampa, FL 21740. American Agricultural Economics Association.*

Durtschi, C., Hillison, W., & Pacini, C. (2004). Effective use of Benford's law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, *V*, 17-34.

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research*, *23*(1), 123-145.

Dye, R. A. (2001). An evaluation of "essays on disclosure" and the disclosure literature in accounting. *Journal of Accounting and Economics*, *32*, 181-235.

Federal Aviation Administration National Wildlife Strike Database. (2016). Online. Retrieved from https://wildlife.faa.gov

Federal Aviation Administration National Wildlife Strike Database. (2018). Online. Retrieved from https://wildlife.faa.gov

Forbes. (2016). *How M&A has driven the consolidation of the US airline industry over the last decade.* Online. Retrieved from https://www.forbes.com/sites/greatspeculations/2016/05/04/how-ma-has-driven-the-consolidation-of-the-us-airline-industry-over-the-last-decade/#4151315b2bba.

Gelb, D. S., Henry, T. F., & Holtzman, M. P. (2008). Deregulation and voluntary disclosure by the airlines: A case study. *Research in Accounting Regulation*, *20*, 89-102.

Gigler, F. (1994). Self-enforcing of voluntary disclosures. *Journal of Accounting Research*, *32*(2), 224-240.

Golbe, D. L. (1986). Safety and profits in the airline industry. *Journal of Industrial Economics*, *34*(3), 305-318.

Harris, M. S. (1998). The association between competition and manager's business segment reporting decisions. *Journal of Accounting Research*, *36*(1), 111-128.

Healy, P. M., & Palepu, K. G. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*, 405-440.

Hill, T. P. (1995). A statistical derivation of the significant-digit law. *Statistical Science*, *10*(4), 354-363.

Hodson, L. (2017). In Person Interview.

IATA. (2010). *IATA economic briefing airline fuel and labour cost share* (Tech. Rep.). Miami: International Air Transport Association

Komarek, P., & Moore, A. W. (2005). Making logistic regression a core data mining tool. In *Research showcase at Carnegie Mellon University.* Online. Retrieved from https://library.cmu.edu/datapub/sc/opencmu/research-showcase-cmu

Lakshminarayan, K., Harp, S. A., Goldman, R., & Samad, T. (1996). Imputation of missing data using machine learning techniques. In *KDD-96 proceedings.* Online. Retrieved from https://www.aaai.org/Press/Proceedings/kdd96.php.

Livernois, J., & McKenna, C. (1999). Truth or consequences: Enforcing pollution standards with self-reporting. *Journal of Public Economics*, *71*, 415-440.

Lohatepanot, M., & Barnhart, C. (2004). Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, *38*(1), 19-32.

Madsen, P. M. (2013). Perils and profits: A reexamination of the link between profitability and safety in U.S. aviation. *Journal of Management*, *39*(3).

McFadden, D. (1974). *Frontiers in econometrics, economic theory and mathematical economics*. Cambridge: Academic Press.

Morrow, J. (2010). *Benford's law, families of distributions and a test basis* (Tech. Rep.). Madison: University of Wisconsin-Madison.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87-106.

Newman, P., & Sansing, R. (1993). Disclosure policies with multiple users. *Journal of Accounting Research*, *31*(1), 92-112.

Nigrini, M. (1996, Spring). A taxpayer compliance application of Benford's law. *The Journal of the American Taxation Association*, *18*(1), 72-91.

Peterman, C. A. (2014). The future of airline mergers after the US Airways and American Airlines merger. *Journal of Air Law and Commerce*, *79*, 781-815.

Raghavan, S., & Rhoades, D. L. (2005). Revisiting the relationship between profitability and aircraft safety in the US airline industry. *Journal of Air Transport Management*, *11*, 283-290.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. *arXiv preprint*, *1710*(05941).

Rauch, B., Goettsche, M., Mouaaouy, F. E., & Geidel, F. (2013). *Empirical methods in competition analysis–applying Benford's law to the Western Australian petroleum market* (Tech. Rep.). Catholic University of Eichstaett-Ingolstadt. Online. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2364384.

Rhoades, S. A. (1993). The Herfindahl-Hirschman Index. *Federal Reserve Bulletin*, *79*(3), 188-189.

Rose, N. L. (1990). Profitability and product quality: Economic determinants of airline safety performance. *Journal of Political Economy*, *98*(5), 944-964.

Rose, N. L. (1992). Fear of flying? Economic analyses of airline safety. *Journal of Economic Perspectives*, *6*(2), 75-94.

Shehata, N. F. (2014). Theories and determinants of voluntary disclosure. *Accounting and Finance Research*, *3*(1), 18-26.

Shen, Y. (2017). Market competition and market price: Evidence from the United/Continental airline merger. *Economics of Transportation*, *10*, 1-7.

Silberberg, E. (1978). *The structure of economics: A mathematical analysis*. New York: McGraw-Hill.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*, 557-586.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Andrews DWK identification and inference for econometric models* (p. 80-108). New York: Cambridge University Press.

Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition*. Cambridge: Academic Press.

Tirole, J. (1988). *The theory of industrial organization*. Cambridge: The MIT Press.

Varian, H. (1972, June). Benford's law–letters to the editor. *American Statistician*, *26*(3), 62-66.

Varian, H. (1992). *Microeconomic analysis* (Third ed.). New York: W.W. Norton and Company, Inc.

Verlinden, B., Duflo, J., P.Collin, & Cattrysse, D. (2008). Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. *International Journal of Production Economics*, *111*, 484-492.

Verrecchia, R. E. (1983). Discretionary disclosure. *Journal of Accounting and Economics*, *5*, 179-194.

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, *32*, 97-180.

Verrecchia, R. E., & Weber, J. (2006). Redacted disclosure. *Journal of Accounting Research*, *44*(4), 791-814.

Wenning, K. M., Begier, M. J., & Dolbeer, R. A. (2004). Wildlife hazard management at airports: Fifteen years of growth and progress for wildlife services. In R. Timm & W. Gorenzel (Eds.), *Proceedings of 21st pest conference* (p. 295-301). University of California Davis. Online. Retrieved from http://digitalcommons.unl.edu/vpccollection/.

Zahran, S., Iverson, T., Weiler, S., & Underwood, A. (2014). Evidence that the accuracy of self-reported led emissions data improved: A puzzle and discussion. *Journal of Risk and Uncertainty*, *49*, 235-257.

Appendix A

Cost-minimization

Consider the following objective function $\mathcal{L}$ for a representative commercial airline that seeks to minimize costs for a given level of output, y:

$$\min_{d,x} \mathcal{L} = \delta d + \gamma x + \lambda[y - f(d,x)] \tag{A.1}$$

Where d and x represent the firm's choice of inputs: the disclosure of non-performance information and a non-disclosure input representative of all other factors relevant to production, respectively. δ and γ are each input's associated exogenously given factor price. $f(d,x)$ is the airline's quasiconcave production function, which is increasing in both inputs[31], and λ is the Lagrange multiplier associated with the output constraint y. Subsequent minimization of (A.1) requires first order necessary conditions where the leading subscript associated with f notates the first derivative of the production function with respect to that variable:

$$\frac{\partial \mathcal{L}}{\partial d} : \delta - \lambda f_d = 0 \tag{A.2}$$

$$\frac{\partial \mathcal{L}}{\partial x} : \gamma - \lambda f_x = 0 \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} : y - f(d,x) = 0 \tag{A.4}$$

Because the interest of this analysis is the effect of profitability on disclosure, comparative statics can be used to investigate the impact of exogenous change in a *profit-influencing* parameter on the airline's optimal choice of d (Golbe, 1986). In the spirit of Dedman and Lennox (2009)'s hypothesis that firms choose to limit their choice of disclosure to hide sources of comparative advantage from competitors (e.g., low costs), consider an

[31] f_d, $f_x > 0$

exogenous change in γ, the price of the non-disclosure input x. Invoking the implicit function theorem and differentiating (A.2), (A.3), and (A.4) with respect to γ gives the following system of equations in matrix form:

$$\underbrace{\begin{bmatrix} -\lambda f_{dd} & -\lambda f_{dx} & -f_d \\ -\lambda f_{dx} & -\lambda f_{xx} & -f_x \\ -f_d & -f_x & 0 \end{bmatrix}}_{\boldsymbol{\Delta}} \begin{bmatrix} \frac{\partial d}{\partial \gamma} \\ \frac{\partial x}{\partial \gamma} \\ \frac{\partial \lambda}{\gamma} \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} \tag{A.5}$$

Where the latter subscript is indicative of the indexed variable's second derivative with respect to the input indicated by the subscript. Applying Cramer's rule to solve for the effect of an exogenous change in profit (via a change in γ) on the airline's optimal choice of strategic disclosure, $\frac{\partial d}{\partial \gamma}$ implies:

$$\frac{\partial d}{\partial \gamma} = -\frac{\begin{vmatrix} 0 & -\lambda f_{dx} & -f_d \\ -1 & -\lambda f_{xx} & -f_x \\ 0 & -f_x & 0 \end{vmatrix}}{\begin{vmatrix} \Delta \end{vmatrix}} \tag{A.6a}$$

Which may be rewritten:

$$\frac{\partial d}{\partial \gamma} = \frac{-(f_d f_x)}{\begin{vmatrix} \Delta \end{vmatrix}} \tag{A.6b}$$

In this simple two-input context, (A.6b) can be easily signed. f_d and f_x are positive by assumption, and the sign of the determinant of Δ must be less than zero as a sufficient condition for cost-minimization. Thus, this framework predicts $\frac{\partial d}{\partial \gamma}$ to be positive, implying an increase in profits via a decrease in non-disclosure input costs would lower firms' choice of d. The intuition behind this result is straightforward: if more profitable firms face a (relatively) higher cost of disclosure, the relationship between profitability and disclosure is negative. Assuming the disclosure of information carries a benefit in terms of production and is simultaneously costly for the firm, increases in profits due to decreases in prices of non-disclosure inputs would cause firms to shift resources away from disclosure and toward less expensive factors of production.

Appendix B

Summary statistics

Table B.1: Summary statistics

Statistic	Mean	St. Dev.	Min	Max
Dependent variable				
Cost reported (1=Yes, 0=No)	0.185	0.388	0	1
Mkt., oper., and fin. characteristics				
Herfindahl-Hirschman index	12.483	1.586	10.786	16.527
Net operating income (USD billions)	0.151	2.245	−8.314	7.802
Load factor (pct.)	78.208	5.332	65.900	85.500
Four firm concentration ratio (pct.)	63.452	5.886	57.017	78.749
Entropy index	−2.262	0.115	−2.400	−1.990
Average stage length (100's of aircraft miles)	10.463	2.953	5.150	17.200
Daily departures (per aircraft)	4.158	1.282	2.590	7.300
Strike characteristics				
Animal ingested (1=Yes, 0=No)	0.416	0.493	0	1
Component damaged (1=Damaged, 0=Otherwise)				
Radome	0.178	0.382	0	1
Windshield	0.019	0.136	0	1
Nose	0.062	0.241	0	1
Engine	0.418	0.493	0	1
Wing/rotor	0.175	0.380	0	1
Fuselage	0.053	0.224	0	1
Landing gear	0.043	0.202	0	1
Tail	0.050	0.218	0	1

Lights	0.062	0.241	0	1
Other	0.094	0.292	0	1
Effect on flight				
Unknown	0.142	0.349	0	1
Aborted take-off	0.037	0.188	0	1
Engine shut down	0.037	0.188	0	1
None	0.510	0.500	0	1
Other	0.089	0.285	0	1
Precautionary landing	0.186	0.389	0	1
Size of animal struck				
Large	0.293	0.455	0	1
Medium	0.495	0.500	0	1
Small	0.212	0.409	0	1
FAA region of strike occurrence				
Central (ACE)	0.056	0.230	0	1
Eastern (AEA)	0.234	0.423	0	1
Great Lakes (AGL)	0.133	0.340	0	1
New England (ANE)	0.047	0.213	0	1
Northwest Mountain (ANM)	0.132	0.339	0	1
Southern (ASO)	0.236	0.425	0	1
Southwest (ASW)	0.161	0.367	0	1
Time of day of strike occurrence				
Unknown	0.134	0.341	0	1
Dawn	0.033	0.179	0	1
Day	0.529	0.499	0	1
Dusk	0.039	0.193	0	1
Night	0.265	0.441	0	1

Phase of flight of strike occurrence				
Unknown	0.005	0.069	0	1
Approach	0.415	0.493	0	1
Climb	0.277	0.448	0	1
Descent	0.041	0.198	0	1
Landing Roll	0.080	0.271	0	1
Take-off run	0.180	0.384	0	1
Taxi	0.003	0.058	0	1

Appendix C

Lag structure

Table C.1: Lag structure–average partial effects

	(1)	(2)	(3)	(4)
	$ReportCost_{ijt}$	$ReportCost_{ijt}$	$ReportCost_{ijt}$	$ReportCost_{ijt}$
NOI_{ijt}	0.00448	0.00793	0.00384	0.00786
	(0.00789)	(0.00792)	(0.00830)	(0.00872)
NOI_{ijt-1}		-0.0184**	-0.0113	-0.0118
		(0.00901)	(0.00954)	(0.0103)
NOI_{ijt-2}			-0.0104	-0.0144
			(0.00973)	(0.0104)
NOI_{ijt-3}				0.0000731
				(0.0114)
HHI_{it}	0.0200	0.0220	0.0498**	0.0553***
	(0.0218)	(0.0218)	(0.0235)	(0.0212)
Strike-specific controls	Yes	Yes	Yes	Yes
Month/year FE	Yes	Yes	Yes	Yes
Airline FE/operations controls	Yes	Yes	Yes	Yes
Airport FE	Yes	Yes	Yes	Yes
N	1476	1476	1340	1205

Robust standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Of additional interest in this chapter is determining the proper (if any) lag structure of NOI to include in the estimation of (1.1). Following Conroy, Low, and Weiler (2017), Table C.1 presents results from the method utilized to accomplish this task: beginning with a long[32] lag structure (in column (4)) and successively shortening the lag length until the marginal lag is significant. All columns represent average marginal effects following the estimation of (1.1), including respective lags via OLS.[33] Following this process, column

[32]Three-year

[33]No IV

(2) of Table C.1 represents the chosen lag structure, including one lag of net operating income in addition to the contemporaneous level described in (1.1).[34].

[34]Potential multicollinearity concerns arise for the two measures of profit However, the correlation between NOI_{ijt} and its one year lag is $.44$, and moreover, the VIF associated with the one-year lag of NOI_{ijt} from column (2) of Table C.1 is 2.86.

Appendix D

Training set summary statistics: Features

Table D.1: Training set summary statistics: Features

	Mean	*Std.*	*Min.*	*Max*
AC_CLASS_A	0.741896	0.437593	0.0	1.0
AC_CLASS_B	0.010711	0.102937	0.0	1.0
AC_CLASS_J	0.000031	0.005601	0.0	1.0
AC_CLASS_missing_value	0.247362	0.431480	0.0	1.0
AC_MASS_1.0	0.037967	0.191118	0.0	1.0
AC_MASS_2.0	0.042134	0.200895	0.0	1.0
AC_MASS_3.0	0.137877	0.344772	0.0	1.0
AC_MASS_4.0	0.490325	0.499908	0.0	1.0
AC_MASS_5.0	0.007435	0.085908	0.0	1.0
AC_MASS_missing_value	0.284262	0.451064	0.0	1.0
BIRDS_SEEN_0	0.000031	0.005601	0.0	1.0
BIRDS_SEEN_1	0.213146	0.409532	0.0	1.0
BIRDS_SEEN_11-100	0.029239	0.168477	0.0	1.0
BIRDS_SEEN_2-10	0.118413	0.323098	0.0	1.0
BIRDS_SEEN_OVER 100	0.003878	0.062150	0.0	1.0
BIRDS_SEEN_missing_value	0.635292	0.481350	0.0	1.0
BIRDS_STRUCK_1	0.842515	0.364258	0.0	1.0
BIRDS_STRUCK_11-100	0.006030	0.077418	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	Mean	*Std.*	*Min.*	*Max*
BIRDS_STRUCK_2-10	0.120346	0.325367	0.0	1.0
BIRDS_STRUCK_OVER 100	0.000213	0.014604	0.0	1.0
BIRDS_STRUCK_missing_value	0.030896	0.173036	0.0	1.0
INDICATED_DAMAGE	0.026516	0.160665	0.0	1.0
DAMAGE_A	0.000006	0.002505	0.0	1.0
DAMAGE_B	0.000063	0.007921	0.0	1.0
DAMAGE_C	0.000471	0.021688	0.0	1.0
DAMAGE_D	0.000634	0.025166	0.0	1.0
DAMAGE_E	0.002686	0.051752	0.0	1.0
DAMAGE_M	0.011363	0.105991	0.0	1.0
DAMAGE_M?	0.002792	0.052767	0.0	1.0
DAMAGE_N	0.678210	0.467164	0.0	1.0
DAMAGE_NO DATA	0.002974	0.054455	0.0	1.0
DAMAGE_S	0.009067	0.094787	0.0	1.0
DAMAGE_missing_value	0.291735	0.454562	0.0	1.0
DAM_ENG1	0.004204	0.064702	0.0	1.0
DAM_ENG2	0.003307	0.057409	0.0	1.0
DAM_ENG3	0.000307	0.017532	0.0	1.0
DAM_ENG4	0.000144	0.012012	0.0	1.0
DAM_FUSE	0.001788	0.042250	0.0	1.0
DAM_LG	0.002227	0.047144	0.0	1.0
DAM_LGHTS	0.001249	0.035314	0.0	1.0
DAM_NOSE	0.002435	0.049281	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	Mean	*Std.*	*Min.*	*Max*
DAM_OTHER	0.002460	0.049534	0.0	1.0
DAM_PROP	0.001437	0.037879	0.0	1.0
DAM_RAD	0.002780	0.052649	0.0	1.0
DAM_TAIL	0.001487	0.038534	0.0	1.0
DAM_WINDSHLD	0.002460	0.049534	0.0	1.0
DAM_WING_ROT	0.008471	0.091646	0.0	1.0
EFFECT_ABORTED TAKE-OFF	0.011947	0.108647	0.0	1.0
EFFECT_ENGINE SHUT DOWN	0.001487	0.038534	0.0	1.0
EFFECT_NONE	0.507404	0.499947	0.0	1.0
EFFECT_OTHER	0.010190	0.100430	0.0	1.0
EFFECT_PRECAUTIONARY LANDING	0.026453	0.160480	0.0	1.0
EFFECT_missing_value	0.442519	0.496686	0.0	1.0
ENG_1_POS_1	0.359362	0.479815	0.0	1.0
ENG_1_POS_2	0.000025	0.005010	0.0	1.0
ENG_1_POS_3	0.000540	0.023223	0.0	1.0
ENG_1_POS_4	0.071605	0.257834	0.0	1.0
ENG_1_POS_5	0.231951	0.422079	0.0	1.0
ENG_1_POS_6	0.007737	0.087617	0.0	1.0
ENG_1_POS_7	0.031090	0.173563	0.0	1.0
ENG_1_POS_C	0.000056	0.007515	0.0	1.0
ENG_1_POS_missing_value	0.297633	0.457219	0.0	1.0
ENG_2_POS_1.0	0.343023	0.474721	0.0	1.0
ENG_2_POS_2.0	0.000006	0.002505	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	$Mean$	$Std.$	$Min.$	Max
ENG_2_POS_3.0	0.001161	0.034051	0.0	1.0
ENG_2_POS_4.0	0.075150	0.263634	0.0	1.0
ENG_2_POS_5.0	0.202323	0.401733	0.0	1.0
ENG_2_POS_6.0	0.052700	0.223434	0.0	1.0
ENG_2_POS_7.0	0.000006	0.002505	0.0	1.0
ENG_2_POS_missing_value	0.325630	0.468611	0.0	1.0
ENG_3_POS_1	0.029365	0.168828	0.0	1.0
ENG_3_POS_3	0.000013	0.003542	0.0	1.0
ENG_3_POS_4	0.002315	0.048062	0.0	1.0
ENG_3_POS_5	0.031699	0.175198	0.0	1.0
ENG_3_POS_CHANGE CODE	0.000282	0.016801	0.0	1.0
ENG_3_POS_missing_value	0.936326	0.244172	0.0	1.0
ENG_4_POS_1.0	0.011545	0.106827	0.0	1.0
ENG_4_POS_3.0	0.000013	0.003542	0.0	1.0
ENG_4_POS_4.0	0.006946	0.083053	0.0	1.0
ENG_4_POS_5.0	0.000075	0.008677	0.0	1.0
ENG_4_POS_missing_value	0.981421	0.135033	0.0	1.0
INGESTED	0.040935	0.198141	0.0	1.0
NUM_ENGS_1.0	0.036712	0.188055	0.0	1.0
NUM_ENGS_2.0	0.611135	0.487494	0.0	1.0
NUM_ENGS_3.0	0.050102	0.218157	0.0	1.0
NUM_ENGS_4.0	0.018610	0.135145	0.0	1.0
NUM_ENGS_missing_value	0.283440	0.450670	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	Mean	*Std.*	*Min.*	*Max*
PHASE_OF_FLT_APPROACH	0.276846	0.447441	0.0	1.0
PHASE_OF_FLT_ARRIVAL	0.000897	0.029941	0.0	1.0
PHASE_OF_FLT_CLIMB	0.107169	0.309329	0.0	1.0
PHASE_OF_FLT_DEPARTURE	0.002635	0.051268	0.0	1.0
PHASE_OF_FLT_DESCENT	0.015605	0.123941	0.0	1.0
PHASE_OF_FLT_EN ROUTE	0.013735	0.116389	0.0	1.0
PHASE_OF_FLT_LANDING	0.005051	0.070891	0.0	1.0
PHASE_OF_FLT_LANDING ROLL	0.120609	0.325674	0.0	1.0
PHASE_OF_FLT_LOCAL	0.001876	0.043273	0.0	1.0
PHASE_OF_FLT_PARKED	0.000577	0.024019	0.0	1.0
PHASE_OF_FLT_TAKE-OFF RUN	0.126489	0.332400	0.0	1.0
PHASE_OF_FLT_TAXI	0.002259	0.047474	0.0	1.0
PHASE_OF_FLT_missing_value	0.326251	0.468842	0.0	1.0
PRECIP_FOG	0.010673	0.102758	0.0	1.0
PRECIP_FOG, RAIN	0.001324	0.036362	0.0	1.0
PRECIP_FOG, RAIN, SNOW	0.000025	0.005010	0.0	1.0
PRECIP_FOG, SNOW	0.000069	0.008308	0.0	1.0
PRECIP_NONE	0.451786	0.497672	0.0	1.0
PRECIP_RAIN	0.029095	0.168074	0.0	1.0
PRECIP_RAIN, SNOW	0.000082	0.009031	0.0	1.0
PRECIP_SNOW	0.001613	0.040124	0.0	1.0
PRECIP_missing_value	0.505333	0.499973	0.0	1.0
SIZE_LARGE	0.053679	0.225383	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	Mean	*Std.*	*Min.*	*Max*
SIZE_MEDIUM	0.321395	0.467013	0.0	1.0
SIZE_SMALL	0.583815	0.492926	0.0	1.0
SIZE_missing_value	0.041111	0.198547	0.0	1.0
SKY_NO CLOUD	0.248849	0.432347	0.0	1.0
SKY_OVERCAST	0.089971	0.286141	0.0	1.0
SKY_SOME CLOUD	0.171741	0.377156	0.0	1.0
SKY_missing_value	0.489440	0.499890	0.0	1.0
STR_ENG1	0.046369	0.210283	0.0	1.0
STR_ENG2	0.037716	0.190510	0.0	1.0
STR_ENG3	0.002723	0.052113	0.0	1.0
STR_ENG4	0.001738	0.041654	0.0	1.0
STR_FUSE	0.103555	0.304684	0.0	1.0
STR_LG	0.043903	0.204880	0.0	1.0
STR_LGHTS	0.003062	0.055251	0.0	1.0
STR_NOSE	0.123696	0.329236	0.0	1.0
STR_OTHER	0.091715	0.288624	0.0	1.0
STR_PROP	0.018384	0.134337	0.0	1.0
STR_RAD	0.105870	0.307673	0.0	1.0
STR_TAIL	0.008515	0.091881	0.0	1.0
STR_WINDSHLD	0.141510	0.348548	0.0	1.0
STR_WING_ROT	0.106755	0.308803	0.0	1.0
TIME_OF_DAY_DAWN	0.019903	0.139667	0.0	1.0
TIME_OF_DAY_DAY	0.387058	0.487079	0.0	1.0

Continued on next page

Table D.1: Training set summary statistics: Features

	$Mean$	$Std.$	$Min.$	Max
TIME_OF_DAY_DUSK	0.027432	0.163340	0.0	1.0
TIME_OF_DAY_NIGHT	0.181121	0.385120	0.0	1.0
TIME_OF_DAY_UNKNOWN	0.000163	0.012772	0.0	1.0
TIME_OF_DAY_missing_value	0.384322	0.486436	0.0	1.0
TYPE_ENG_A	0.040916	0.198097	0.0	1.0
TYPE_ENG_A/C	0.000358	0.018908	0.0	1.0
TYPE_ENG_B	0.002240	0.047276	0.0	1.0
TYPE_ENG_B/D	0.000339	0.018404	0.0	1.0
TYPE_ENG_C	0.066880	0.249816	0.0	1.0
TYPE_ENG_D	0.595875	0.490723	0.0	1.0
TYPE_ENG_F	0.009600	0.097509	0.0	1.0
TYPE_ENG_X	0.000006	0.002505	0.0	1.0
TYPE_ENG_missing_value	0.283785	0.450835	0.0	1.0
WARNED_N	0.252783	0.434609	0.0	1.0
WARNED_Y	0.174313	0.379380	0.0	1.0
WARNED_missing_value	0.572904	0.494658	0.0	1.0