Factors Influencing Airline Delays of United States

Xiaomeng Zhao

College of General Study, Boston University, Boston, MA, United States

zhaoxm@bu.edu

Abstract. This article analyzes and studies the causes of airline delays originating in the United States. The aim is to find out the most significant causes of flight delays and help airlines better optimize routes and reduce potential losses. Previous studies have shown that the main causes of airplane delays are delays in the national aviation system and the airlines' own delays. This article uses linear regression to analyze the data. The conclusion is that the occurrence of aircraft delays may be related to factors such as National Aviation System, Aircraft Arriving and weather. While analyzing the factors that cause flight delays, the paper also analyzed the interactions between these factors and a strong correlation was found between Air Carrier Delay and Aircraft Arriving Late Delay, this may have contributed to the fact that the effect of Air Carrier Delay on flight delays was not significant enough for the analysis in this paper. In addition, Airline Diverted, Flight Cancellation, and Security Delay had very little effect on with flight delays.

Keywords: Fight delays; delay causes; correlation analysis; linear regression model.

1. Introduction

Airplanes are one of the most common transportation methods used by modern people's daily life. Its quality service and efficiency have led people to trust and continue to choose air travel. It is undeniable that the global use of airplanes has also contributed to global economic development and connections between different cultures. Despite the many advantages of airplanes, civil aviation companies still face many challenges. For example, problems such as environmental pollution, aviation safety and flight delays, etc. [1]. As a result, civil aviation companies are still trying to improve their technology and services in order to provide more sustainable and convenient services.

As a mode of transportation, the reason passengers choose to ride the airline depends largely on the punctuality of the aircraft. For many passengers, airplane delays are a serious problem. Flight delays will reduce customer satisfaction and may affect the civil aviation companies' reputation and lead to customer attrition in the long-term [2]. According to the research about the airline delay probabilities at an airport from Ronald Wesonga, "The average delay probabilities for ground delays and air holds are 0.94 and 0.82, respectively, at a delay threshold of 50 percent [3]. Airplane delays seem to be becoming a frequent occurrence, so the deeper understanding of the airline delay factors is necessary. The factors affecting aircraft delays seem to be summarized in only a few ways. Weather is always considered as a critical impact on the departure of aircraft [4]. In addition, air traffic control has a huge impact on whether a plane can take off on time [5].

At the same time, a deeper appreciation of the causes of aircraft delays will assist airline companies further optimize route planning and prepare alternatives. Aircraft delays can have significant economic and safety impacts [6]. Unexpected delays can cause delays to all other scheduled flights. On the other hand, unplanned delays of connecting flights can lead to serious problems in continuing the journey as well as a significant loss of profit for the airline [7]. A stronger grasp of the causes of airline delay can maximize the avoidance of losses to airlines and airports due to delays and give passengers the best possible experience.

2. Methods

2.1. Data Sources

The data for this literature is collected from the Kaggle website, which was compiled by Eugeniy Osetrov from the United States Department of Transportation website and published in 2021 for most of the United States civil aviation companies Airlines delays date from December 2019 and 2020 [8].

2.2. Variable Selection

The data in this paper count a total of 3352 volumes, which includes different airline companies in different airports in the United States. Including 28 civil aviation companies, which are American Airlines Inc., Alaska Airlines Inc., JetBlue Airways, Delta Air Lines Inc., Frontier Airlines Inc., etc. The x variable the different factors (Air Carrier, Weather, National Aviation System, Security, Aircraft Arriving, Cancelled, and Diverted) contains in the data that may cause the airline delay [9]. The y variable is the slippage rate of the airline, which means the number of flights with flight delays of more than 15 minutes at different airports for each airline in December 2019 divided by the total number of arrival flights in the United States. It can be represented as:

$$y = \frac{\text{Number of flights delayed by more than 15 minutes}}{\text{Total number of flights arriving at the airport}} \quad (1)$$

Table 1. Descriptive Analysis

Factors	Elements	Min	Max	Mean	Std. Deviation	Median
Air Carrier Delay	x1	0	697	16.065	41.76	4.75
Weather Delay	x2	0	89.42	1.443	4.822	0.06
National Aviation System Delay	x3	0	1039.54	16.183	56.423	2.98
Security Delay	x4	0	17.31	0.137	0.646	0
Aircraft Arriving Late	x5	0	819.66	17.166	55.447	3.28
Cancelled	x6	0	224	2.885	10.127	0
Diverted	x7	0	42	0.576	2.098	0

*National Aviation System: e.g., Air traffic is heavy.
**Security: Amount of cancellations of flights because of security breaches.

From the table 1, the maximum value of exceeds the mean by three standard deviations, which indicates that the data fluctuates greatly as shown below.

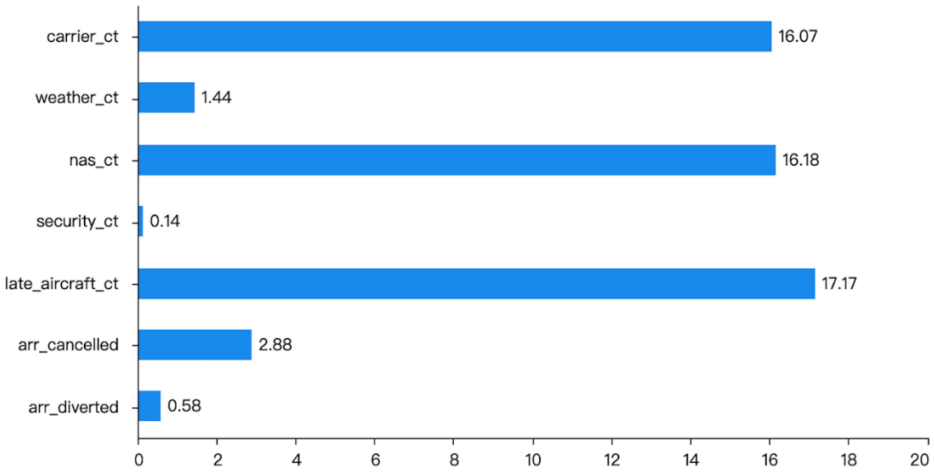

Fig. 1 Bar Graph

From figure 1 (The elements in the y-axis are corresponding to the factors in the Table 1 one by one), Air Carrier, National Aviation System, and Aircraft Arriving Late are the notable factors that cause flights to be delayed for more than fifteen minutes.

2.3. Research Protocol

This paper uses SPSSAU to analyze the relationship between the effect of X on Y, i.e., the relationship between the 7 factors on airplane delay., and the relationship between variables X. Then, this paper uses the Linear Regression model [10, 11], the percentage of airplane delays variable (Y), and the 7 factors are independent or dependent variables (X).

3. Results and Discussion

3.1. F-Test

As can be seen from the table 2, F values is 9.961, and the p value is 0.000 less than 0.05. Thus, the model construction used in this paper is meaningful and at least one x variable will influence y.

Table 2. Anova test results

	Sum of Squares	df	Mean Square	F	p value
Regression	0.745	7	0.106	9.961	0.000
Residual	35.638	3335	0.011		
Total	36.383	3342			

3.2. Correlation Analysis

As shown in the Table 3, utilizing the correlation analysis to study the correlation between slipping rate and each of the 7 x-elements. The Pearson's correlation coefficient can indicate the strength of the correlation between x variable and y variable.

Table 3. Pearson Correlation

Elements	Slippage Rate
Air Carrier Delay	0.088**
Weather Delay	0.041*
National Aviation System Delay	0.113**
Security Delay	0.036*
Aircraft Arriving Late	0.121**
Cancelled	0.061**
Diverted	0.065**

* $p<0.05$ ** $p<0.01$

There is a significant positive correlation between slippage rate and Air Carrier Delay, because the correlation coefficient between the two variables is 0.088 and the significance at the 0.01 level. Slippage rate and Weather Delay has a significant positive correlation since the correlation coefficient value between them is 0.041 and it is significant at 0.05 level. National Aviation System Delay has a significant positive relationship with the slippage rate through the correlation coefficient between them is 0.113 and the significance at 0.01 level. Slippage rate and Security Delay has a significant positive correlation since the correlation coefficient value between them is 0.036 and it is significant at 0.05 level. There is a significant positive correlation between slippage rate and Aircraft Arriving Late Delay, because the correlation coefficient between the two variables is 0.121 and the significance at the 0.01 level. Airline Cancelled has a significant positive relationship with the slippage rate through the correlation coefficient between them is 0.061 and the significance at 0.01 level.

Airline Diverted has a significant positive relationship with the slippage rate through the correlation coefficient between them is 0.065 and the significance at 0.01 level. As a result, all the x variables are showing a significant positive correlation with slippage rate. Based on the correlations between the x- and y-variables, the use of linear regression methods for further correlation analysis was justified.

The correlation between x and x variables is examined by means of an autocorrelation test. If there is a large correlation between the two variables, x variable and x variable may influence each other. Removing one of the two x factors that is too highly correlated may necessary for better selecting the factors contributing to the airline delay problem.

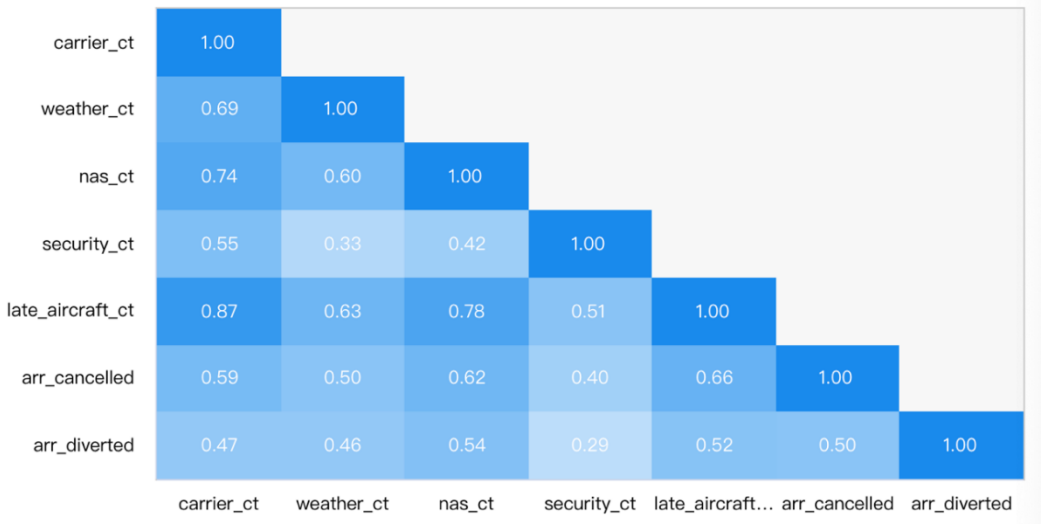

Fig. 2 Correlation Heat Map

From the figure 2 (The elements on the x- and y-axes correspond to the factors in Table 1 one by one.), the darker color means a stronger correlation between the two x variables. Number 0.74 representing a potential strong correlation between National Aviation System Delay and Air Carrier Delay, so there may be one x element that is more apparently associated with the y variable and another x element that is not apparently associated with the airline delay (y variable). Number 0.78 representing a potential strong covariance between two influence factors Aircraft Arriving Late and National Aviation System. Therefore, Aircraft Arriving Late Delay or National Aviation System Delay may be a non-significant factor in the following analysis. Number 0.87 representing a potential strong correlation between Aircraft Arriving Late and Air Carrier Delay, so one x variable may need to be deleted in the following study to get more accurate correlation result between x variable and y variable.

Table 4. Parameter Estimates (n=3343)

	standard coefficients		t	p-value	Covariance Diagnostics	
	Beta	SE			VIF	Tolerance
Constant	16.4	0.2	84.679	0.000**	-	-
Air Carrier Delay	-5	0.1	-1.255	0.21	5.302	0.189
Weather Delay	-5.6	0.1	-2.289	0.022*	2.061	0.485
National Aviation System Delay	7.9	0.1	2.693	0.007**	2.952	0.339
Security Delay	-2.7	0.3	-1.309	0.191	1.475	0.678
Aircraft Arriving Late	17.2	0.1	4.321	0.000**	5.363	0.186
Cancelled	-3.7	0.1	-1.524	0.127	1.973	0.507
Diverted	0.9	0.1	0.417	0.677	1.566	0.639

From the Table 4, it can be seen that all 7 x variables (Air Carrier Problem Delay, Weather Effect Delay, National Aviation System Delay, Security Delay, Aircraft Arriving Late, Flight Cancelled,

and Flight Diverted) are taken as independent variables, while the slippage rate as the dependent variable for linear regression analysis. The model equation is: Slippage rate = 16.4-0.1*Air Carrier Delay-0.1*Weather Delay+0.1*National Aviation System Delay-0.3* Security Delay + 0.1*Aircraft Arriving Late-0.1*cancelled + 0.1*diverted.

Based on the analysis of covariance between x variable and x variable in figure 2 is concluded that there may be a strong correlation between several x variables. According to the Table 4, the factor Air Carrier Delay's VIF value is 5.302 greater than 5 and less than 10. The factor Aircraft Arriving Late Delay's VIF value is 5.363 greater than 5 and less than 10. Further confirmed that there is a high probability of covariance between the x variables, and it may be necessary to consider eliminating the closely correlated independent variables to obtain a more accurate correlation analysis of flight delays.

According to the Table 4, there are three factors will strongly influence on slippage rate. The value of Beta for National Aviation System Delay is 7.9, t=2.693, p=0.007 less than 0.01. The Beta value (regression coefficient value) for Aircraft Arriving Late is 17.2, t=4.321, and p=0.000 less than 0.01. Thus, the National Aviation System and Aircraft Arriving Late factors will have a substantial positive influence on slippage rate. The Weather Delay will have a significant negative effect on slipping rate. This is because the Beta value of Weather Delay is -5.6, t=-2.289, p=0.022 less than 0.05.

Nevertheless, the other 4 potential effect factors including Air Carrier Delay, Air Carrier Delay, Cancelled, and Diverted do not show a significant impact on the airline slipping rate. As shown in the Table 4 the Air Carrier Delay's Beta value is -5, t=-1.255, and p=0.21 more than 0.05. The Beta value of Security Delay is -2.7, t=-1.309, and p=0.191 more than 0.05. The Beta value of Cancelled is -3.7, t=-1.524, and p=0.127 more than 0.05. The value of Beta of Diverted is 0.9, t=0.417, and p=0.677 more than 0.05.

4. Conclusion

This paper mainly focusing on analyzing the 7 potential factors may influence the airline delay, and a variety of data were selected to help analyzing the influence factors. The conclusion of the research shows that the occurrence of delays of airplanes may strongly be related to factors such as National Aviation System, Aircraft Arriving, and Weather. Undoubtedly, since the data comes from the last month of 2019, the results may not be exactly. What's more, the data are used in this paper affected by the pandemic cannot be ignore. The sharp decline in air travel, health and safety protocols and other unpredictable factors the pandemic brought to the airline all may influence the error of the result. Notwithstanding, this study still has some value and advantages. First of all, the data volume for this study totaled 3,352, and the data were complete and not missing, which made the analysis of this study more reliable. Secondly, this experiment also considered autocorrelation when considering the relationship between x-variables and y-variables, i.e., the relationship between x-variables and x-variables. It makes the experiment more precise. Last but not least, it also has a beneficial effect on the improvement of airline' planning of flight operations, which helping them to minimize the problem of airline delay problem and to come up with more well-prepared backup plan. It should be noted that the current study is only based on data for flights at the airport within the United States. There are different climatic and weather conditions at different airports, as well as different transportation constraints caused by factors such as geographic location, seasonality and the occurrence of unusual events. Therefore, further analysis is needed to consider the factors contributing to flight delays based on the combined conditions at local airports.

References

[1] Kireev T, et al. Analysis of the Influence of Factors on Flight Delays in the United States Using the Construction of a Mathematical Model and Regression Analysis. 2022 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2022, 1-5.

[2] Wang F, et al. A flight delay prediction model based on classification regression decision tree algorithm. Journal of Civil Aviation University of China, 2022, 40: 6.

[3] Wesonga Ronald, et al. Parameterized Framework for the Analysis of Probabilities of Aircraft Delay at an Airport. Journal of Air Transport Management, 2012, 1-4.

[4] Markovic D, et al. A Statistical Study of the Weather Impact on Punctuality at Frankfurt Airport. Meteorological Applications, 2008, 15: 293–303.

[5] Zámková Martina, et al. Factors Influencing Flight Delays of a European Airline. Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 2017, 65: 1799-1807.

[6] Ostroumov I, et al. A Probability Estimation of Aircraft Departures and Arrivals Delays. ICCSA 2021. Lecture Notes in Computer Science, 2021, 12950.

[7] Cook Andrew, et al. The Hidden Cost of Airline Unpunctuality. Journal of Transport Economics and Policy (JTEP), 2023, 46: 157-173,

[8] Leonie Hutter, Florian Jaehn, Simone Neumann. Influencing factors on airplane boarding times. Omega, 2019, 177-190.

[9] Nena Adrienne, Lucy Budd, Stephen Ison. Grounded aircraft: An airfield operations perspective of the challenges of resuming flights post COVID. Journal of Air Transport Management, 2020, 89.

[10] Jules Yimga. The airline on-time performance impacts of the COVID-19 pandemic. Transportation Research Interdisciplinary Perspectives, 2021, 10.

[11] Yuan K, Vernachio B. The Impact of Covid-19 on Flight Departure Delay Time and Cancellations at Philadelphia International Airport. Journal of Student Research, 2021.