

# Final Project: Airline Baggage Data

Ian Strawn

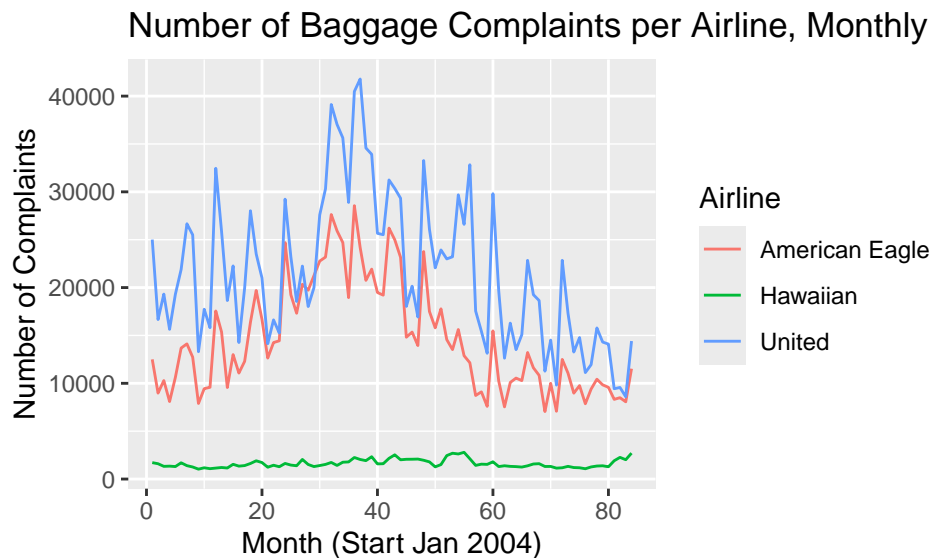
2026-03-18

## Introduction

Air travel is one of the most common forms of transportation, garnering an estimated 2.5 million travelers *per day* in the United States alone. With so much air traffic, it's oftentimes logistically difficult for airlines to keep track of everybody's luggage, and mistakes are often made as a result. If you've never had to experience lost or misplaced luggage during travel, consider yourself lucky; it often involves standing in long lines and waiting a frustrating amount of time to get help getting your stuff back. Looking at *how often* baggage gets lost is a useful goal, as it might be able to shed some light on whether or not certain airlines are better at handling baggage than others, and also might give us some insight as to what times of year might be the worst.

Attempting to model the volume of baggage complaints is inherently going to be seasonal, as travel volume tends to fluctuate throughout the year, and is similar for the same months of different years. Holidays, particularly around Thanksgiving and Christmas, result in higher numbers of people attempting to travel, indicating that air travel (and the number of baggage complaints that go with said travel) likely has some seasonal components to it. As such, I'll operate using a 12-month seasonal cycle with my subsequent analyses.

To take a look at the amount of complaints certain airlines received, I've chosen to pull a dataset from kaggle (listed later in the paper) with monthly data from 3 airlines, taken from 2004-2010. The 3 airlines listed are American Eagle (a subsidiary of American Airlines), Hawaiian Airlines, and United Airlines. Among the variables recorded, the number of baggage complaints per month is included. The time series data for the 3 airlines is as follows:



As a result of taking a look at this data, a useful research question emerges: **is there a difference in the patterns of the number of baggage complaints between airlines?** To answer this, I intend to fit a

variety of models to each of the time series' pictured above, and then compare those models to see if there are any differences between the best-fitting models for each of the airlines. Important to note in the graph is that the 3 airlines clearly have different levels of traffic, so some are inherently going to have more than others (a graph with only the Hawaiian airlines data is included in the **appendix** so that the variation can be seen more clearly). This should not be a problem for the analysis, however; since I'm mostly worried about the model behavior of each airline rather than the frequency of the data, I'll just focus on differences between the models.

## Methods

To analyze these data, I aim to fit various SARIMA models to the different airlines, and then compare those using internal AIC and BIC metrics to determine which models perform the best. ARIMA models are usually used to determine if there are any autoregressive (AR) or moving average (MA) components in a model while also accounting for any differencing that might be necessary to ensure the data is stationary. However, I believe that a SARIMA model (or an ARIMA model with a seasonal component) would be particularly useful here, as flight data is inherently seasonal, which is especially apparent when looking at a plot of the differenced data. In order for a SARIMA model to be used, a certain number of conditions must be met:

- Stationarity: the data expects the mean and variance of the data set to remain constant over time
- Linearity: relationships between observations are linear
- Residuals are IID Normal
- Constant Variance
- Plot diagnostics appear reasonable

Without differencing the data, the stationarity is in jeopardy; from the original plots shown in the introduction, we can see that there appears to be an almost parabolic trend in the data, where it gradually increases, peaks, and then gradually decreases as time goes by. When we difference this data using  $h = 1$ , that trend disappears, and we are left with stationary data (plots of the differenced data sets appear in the **Appendix**). The other assumptions appear to be reasonably met as well - the linearity and constant variance conditions appear to be fine throughout the differenced data, and the residuals being IID is something that we can look at as we fit the models themselves. The only one that might come into question is the validity of the diagnostics of each model; those are going to be model-specific, and I will discuss those in the results section.

## Results

I settled on testing 5 different models for each of the different airlines, and how I came to choose those models will be mentioned in the discussion section:

- $SARIMA(1, 1, 1)(1, 1, 1)_{12}$
- $SARIMA(2, 1, 2)(1, 1, 1)_{12}$
- $SARIMA(3, 1, 1)(1, 1, 1)_{12}$
- $SARIMA(1, 1, 3)(1, 1, 1)_{12}$
- $SARIMA(3, 1, 3)(1, 1, 1)_{12}$

Below are a couple of tables for the AIC and BIC values from each of the models tested, and the airlines that they correspond to. Note that the coefficients of the models tested (e.g. (1,1,1) for a  $SARIMA(1, 1, 1)(1, 1, 1)_{12}$  model) just represent the initial 3 coefficients, not the seasonal ones:

Table 1: AIC Values Across Tested Models

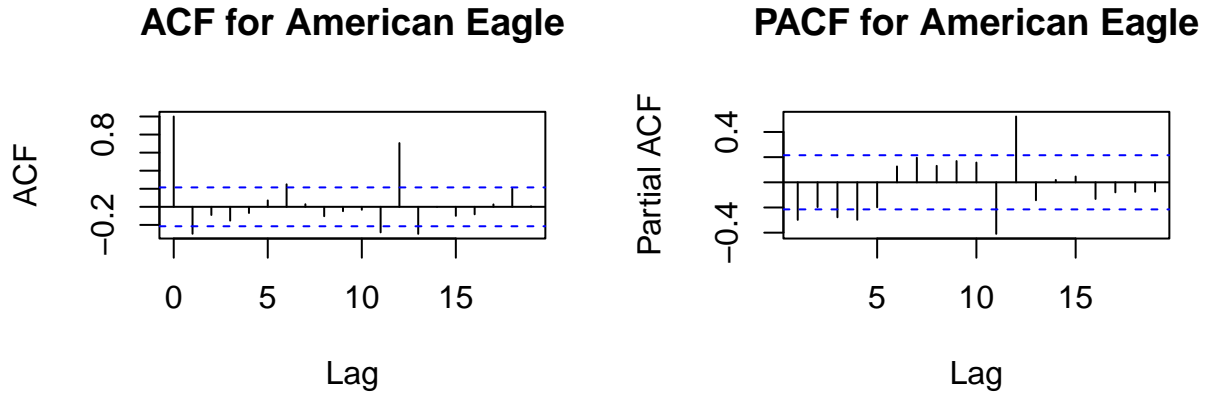
	(1,1,1)	(2,1,2)	(3,1,1)	(1,1,3)	(3,1,3)
AE	18.304	18.334	18.235	18.244	18.289
Hawaiian	14.577	14.620	14.632	14.632	14.637
United	19.319	19.319	19.296	19.319	19.329

Table 2: BIC Values Across Tested Models

	(1,1,1)	(2,1,2)	(3,1,1)	(1,1,3)	(3,1,3)
AE	18.464	18.557	18.459	18.467	18.575
Hawaiian	14.736	14.843	14.855	14.855	14.923
United	19.484	19.542	19.519	19.542	19.616

## Discussion

My general approach for determining which models to evaluate was as follows. Before I had begun creating models, I had determined that differencing the data once was sufficient to help create stationarity, so I decided to not mess with my value of  $h = 1$ , and just kept that going in all of the models that I used. Also, since the seasonal component was a little bit trickier for me to diagnose, I decided that a seasonal component of  $(1, 1, 1)$  in the SARIMA model would be enough to work with (this was after poking around with other values and settling on that simple number). Then, I took a look at the ACF and PACF of each of the airlines' differenced datasets, and used those to inform my decisions as to which SARIMA models would be most useful to try. For example, here is the ACF and PACF of the American Eagle differenced data:



Even with the differenced data, there is an obvious seasonal component, illustrated by the significant peaks roughly at round Lag = 12. However, we can also see some other significant spikes, namely around lag = 3 for the PACF in particular. Using the information that I gained from looking at the ACF and PACF plots, I tried to pick a variety of models that balanced simplicity and the information gleaned from plots similar to the ones above.

Ultimately, based on the AIC/BIC values and also looking at the diagnostic plots for each of the models, it seems like the  $SARIMA(3, 1, 1)(1, 1, 1)_{12}$  model performed the best for both American Eagle and United, but the simpler  $SARIMA(1, 1, 1)(1, 1, 1)_{12}$  performed the best for Hawaiian airlines.

American Eagle and United behaved similarly; the  $(3,1,1)$  model showed normal residuals and insignificant Ljung-Box p-values for both airlines, and performed the best in terms of AIC and BIC out of all of the

models considered. Interestingly, some of those models failed some of the diagnostic checks - the (1,1,1) and (2,1,2) models both had a sizeable number of Ljung-Box p-values, indicating that they were poor fits to begin with. Hawaiian airlines performed completely differently - if you take a look at the ACF and PACF plots corresponding to Hawaiian airlines, there are *no significant lags*, meaning that trying to fit an ARIMA or SARIMA model might not even be useful for us to try in the first place! Finally, in pretty much all cases, the residuals behaved approximately normally and looked to be mostly independent of each other.

## Conclusion

To answer the original research question, it does seem like there is a difference in the autocorrelation structures of the data between the airlines - Hawaiian airline baggage complaints appear to behave differently than those from United or American Eagle. There are a variety of reasons why this might be the case. First, and most notably, Hawaiian airlines has noticeably less flights overall than United and American do. While I don't think this is too much of an issue in the way that I did my above analysis, it would still be useful to take into consideration going forwards, so if I were to do this analysis again, I'd consider implementing the overall number of flights somehow (perhaps by looking into the rate of complaints per flight). Another reason that the patterns may be different could just be down to the airlines themselves; while I'm not terribly familiar with the reputation of Hawaiian airlines, American and United have certainly had their share of negative press, so the airlines could just truly have quality differences (and I've got some STRONG feelings about American, but that's beyond the scope of this project). Ultimately, while this is a useful start to analyzing how baggage complaints change over time, there are some other steps that might be useful to try in the future to get deeper into the analysis.

## Source

The data set can be found at the following website:

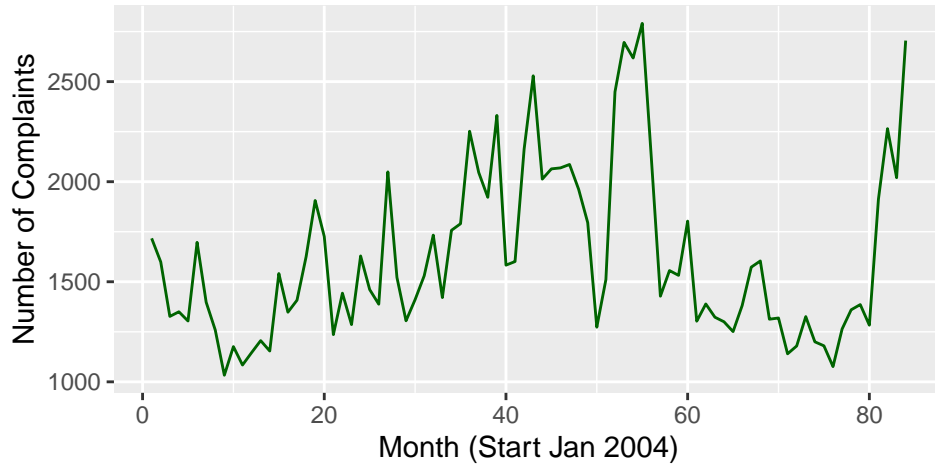
<https://www.kaggle.com/datasets/gabrielsantello/airline-baggage-complaints-time-series-dataset>

# Appendix

## Additional Plots

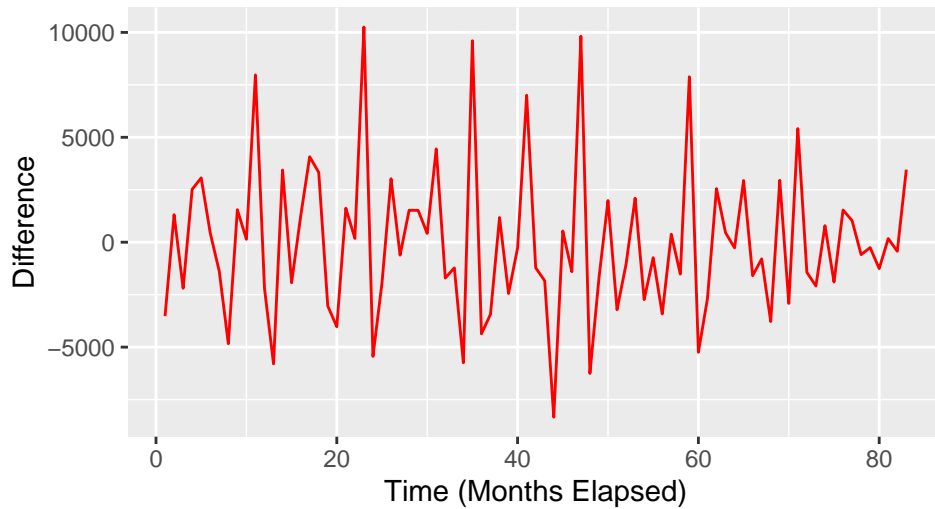
### Additional Plot for Hawaiian Airlines

Number of Baggage Complaints for Hawaiian Airlines  
Monthly

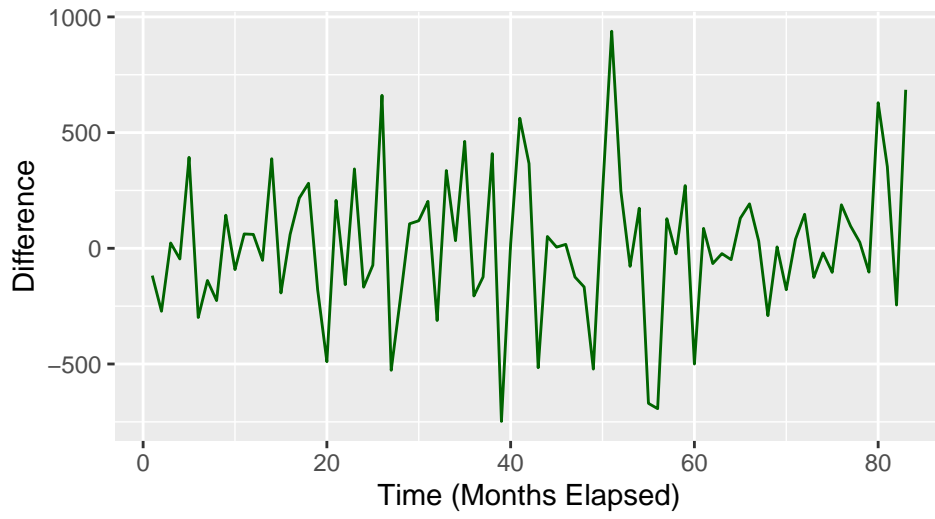


### Plots For Differenced Data

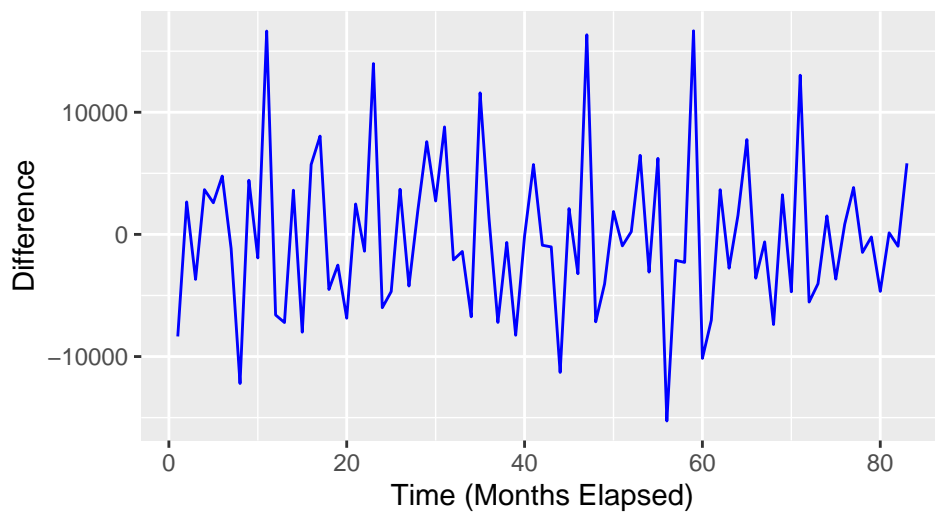
Differenced Data for American Eagle,  $h = 1$



Differenced Data for Hawaiian Airlines,  $h = 1$

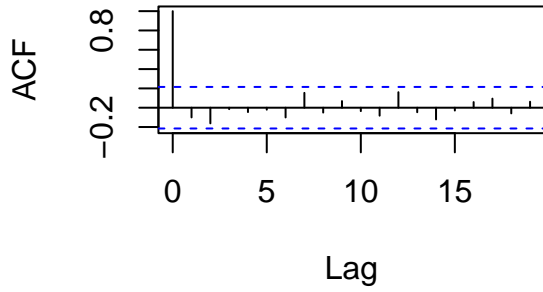


Differenced Data for United Airlines,  $h = 1$

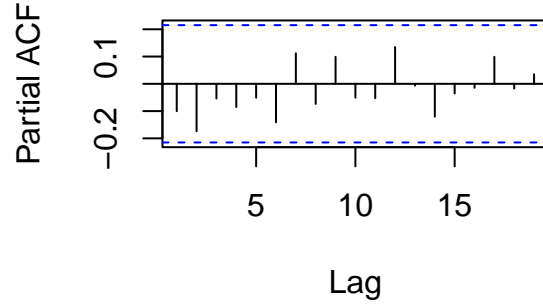


Hawaiian ACF and PACF (from differenced data)

**ACF for Hawaiian**

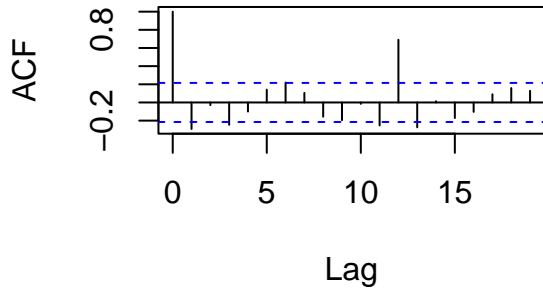


**PACF for Hawaiian**

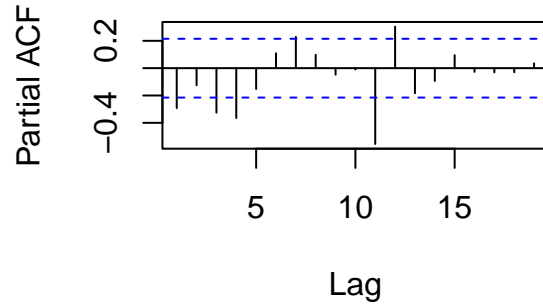


United ACF and PACF (from differenced data)

**ACF for United**



**PACF for United**



## Analysis Code

### American Eagle

```
#Taking a look at the data
#American Eagle, original vs. Differenced data
plot(ae_time, main = "American Eagle Baggage Complaint Data")

#Differencing and plotting
ae_diff = diff(ae_time, lag = 1)
plot(ae_diff, main= "American Eagle Differenced Data")

#Looking at ACF and PACF of differenced data to make sure that
#things look stationary
acf(ae_diff)
pacf(ae_diff)

#Messing around with some models, seasonal differencing only
ae_111 = sarima(ae_time, 1, 1, 1, 1, 1, 1, S = 12)
ae_212 = sarima(ae_time, 2, 1, 2, 1, 1, 1, S = 12)
ae_311 = sarima(ae_time, 3, 1, 1, 1, 1, 1, S = 12)
ae_113 = sarima(ae_time, 1, 1, 3, 1, 1, 1, S = 12)
ae_313 = sarima(ae_time, 3, 1, 3, 1, 1, 1, S = 12)

#Getting AIC and BIC Values for each of the models:
ae_111$ICs
ae_212$ICs
ae_311$ICs
ae_113$ICs
ae_313$ICs
```

### Hawaiian

```
#Hawaiian, original vs. differenced data
plot(hawaiian_time)
hawaiian_diff = diff(hawaiian_time, 1)
plot(hawaiian_diff)

#ACF and PACF
acf(hawaiian_diff)
pacf(hawaiian_diff)

#Messing around with some models, seasonal differencing only
haw_111 = sarima(hawaiian_time, 1, 1, 1, 1, 1, 1, S = 12)
haw_212 = sarima(hawaiian_time, 2, 1, 2, 1, 1, 1, S = 12)
haw_311 = sarima(hawaiian_time, 3, 1, 1, 1, 1, 1, S = 12)
haw_113 = sarima(hawaiian_time, 1, 1, 3, 1, 1, 1, S = 12)
haw_313 = sarima(hawaiian_time, 3, 1, 3, 1, 1, 1, S = 12)

#Getting AIC and BIC Values for each of the models:
haw_111$ICs
haw_212$ICs
haw_311$ICs
haw_113$ICs
```

```
haw_313$ICs
```

## United

```
#United, original vs. differenced
plot(united_time)
united_diff = diff(united_time, 1)
plot(united_diff)

#ACF and PACF
acf(united_diff)
pacf(united_diff)

#Messing around with some models, seasonal differencing only
un_111 = sarima(united_time, 1, 1, 1, 1, 1, 1, S = 12)
un_212 = sarima(united_time, 2, 1, 2, 1, 1, 1, S = 12)
un_311 = sarima(united_time, 3, 1, 1, 1, 1, 1, S = 12)
un_113 = sarima(united_time, 1, 1, 3, 1, 1, 1, S = 12)
un_313 = sarima(united_time, 3, 1, 3, 1, 1, 1, S = 12)

#Getting AIC and BIC Values for each of the models:
un_111$ICs
un_212$ICs
un_311$ICs
un_113$ICs
un_313$ICs
```