Understanding the Impact COVID-19 on American Airlines Stock Prices

Lukas Barrett, Aravind Kotikelapudi

June 7, 2023

Abstract

1 Introduction

1.1 What is the Problem Being Considered?

The COVID-19 pandemic caused the valuation of airlines within the global aviation sector to plummet. Travel was heavily restricted across most of the world with the exception of air cargo carriers. Families and friends struggled to meet with one another in person for momentous celebrations, holidays, and other festivities. It is important that we analyze the underlying trends, seasonal patterns and future value of the aviation industry to ensure the stabilization and full recovery of the economy after the COVID-19 pandemic.

The purpose of this project is to understand the underlying patterns of the publicly traded company American Airlines. By analyzing the closing stock price within the last year, we aim to answer the following questions in order to extrapolate to the broader aviation sector and spark more interest and inquiry around the subject.

Motivating Questions:

1. Is there a rise in American Airlines stock price throughout the holiday season?
2. Can we predict if and when the American Airlines stock price will recover after the COVID-19 pandemic?

1.2 What is the proposed solution(how does this relate to existing literature)?

This research is similar to existing literature in the field of financial time series analysis and forecasting. However, there is limited research in the field discussing the aviation sector and the general recovery of airline stock prices after the COVID-19 pandemic.

1.3 How is the rest of the write-up organized?

In order to answer the following questions, we will perform a trend and seasonality analysis, explore the stationarity(or non-stationarity) of the data, fit a stationary auto-regressive moving average(ARMA) model to the data to forecast future stock prices based on historical data, and finally perform a spectral analysis to understand the underlying periodicities, cyclic patterns, and dominant frequencies that can influence the price movements of the American Airlines stock price.

2 Data Description

2.1 A detailed introduction to the time series data should be given

2.2 What is it that you are looking at? Where did it come from?

The dataset used in this study is the closing price of the publicly traded company American Airlines between May 2022 to May 2023. This dataset was found and extracted from Yahoo! Finance, a trusted source for financial reporting, news, and financial tooling including but not limited to "stock quotes, press releases, financial reports, and original content".

2.3 Data preparation

We find that there is only one outlier in the month of January 2023. The closing price of the American Airlines stock has been tightly bound between $11 and 18$ from May 2022 to May 2023. Additionally, there are no missing values within the dataset. In order to apply the appropriate models, it was imperative that we convert the Date column to a "Date" object that was suitable for time series applications

2.4 What are the interesting features? - Visualizations(See code)

The mean and median of the closing price in the last year are approximately equal (mean = 14.36, median = 14.07) indicating that the distribution follows an approximately normal shape.

The following histogram displays the frequency of times the American Airlines closing stock price is between $11 and 18$ in the last year

The following set of box plots displays the distribution of data based on a five number summary of the American Airlines closing stock price

The ACF plot indicates that the trends are non-stationary, as the majority of lags exceed the dashed blue line representing the significance threshold. The ACF plot is considered non-stationary because the decay of the lags(although exponentially decreasing) is slow relative to stationary time series that experiences much sharper decays.

Because of the gradual decay in the ACF plot and only 2 significant lags in the PACF plot, we conclude that this is an AR process.

3 Analysis

3.1 Initial Data Transformations

The time series of the initial AAL data is shown 1:

3.2 The Smooth Component

First we will be analyzing the trend and seasonality components of the time series.

3.2.1 Trend

The first method to smooth the time series is using a moving average filter: We can see that a filter size of 10 gives the best trend estimation without over-fitting the seasonality(cite lecture notes) [6]

$$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t+j} \tag{1}$$

Subtracting the trend: Due to the moving average filter, we need to get rid of 4 points: 2 at the beginning and 2 at the end.

The second method is differencing, which uses the backshift operator

$$\nabla X_t = (1-B)X_t = X_t - X_{t-1}$$

And by repeated operation, we can get a stationary process with the degree of the estimated polynomial. The mean is $p!b_p$ [6]

$$\nabla^p X_t = p!b_p + \nabla^p Y_t \tag{2}$$

3.2.2 Seasonality

Cannot use small trend method: clearly a non-neglibible trend Using a moving average filter with q=d/2 where d is the period [6]. There are about 5 cycles in a dataset of length 250, so d=50, N=5, n=250. Therefore we can first apply a moving average filter to get $\hat{m}_t = W_t$ before estimating seasonality.

Next, we can estimate the seasonality component with:

$$\mu_k = \frac{1}{N-1}(\sum_{j=2}^{N} x_{k+d(j-1)} - \hat{m}_{k+d(j-1)})$$

$$\mu_k = \frac{1}{N-1}(\sum_{j=2}^{N} x_{k+d(j-1)} - \hat{m}_{k+d(j-1)})$$

Where the first equation applies to ($k = 1, ..., q$) and the second applies to ($k = q+1, ..., d$).

And finally:

$$\hat{s}_k = \mu_k - \frac{1}{d}\sum_{\ell=1}^{d} \mu_\ell \tag{3}$$

Both estimates and the residual plot is shown below 1

NOTE: I used the same moving average filter for the trend estimate as I did for the seasonality estimate. This might cause problems as usually the trend moving average filter are found separately [6].

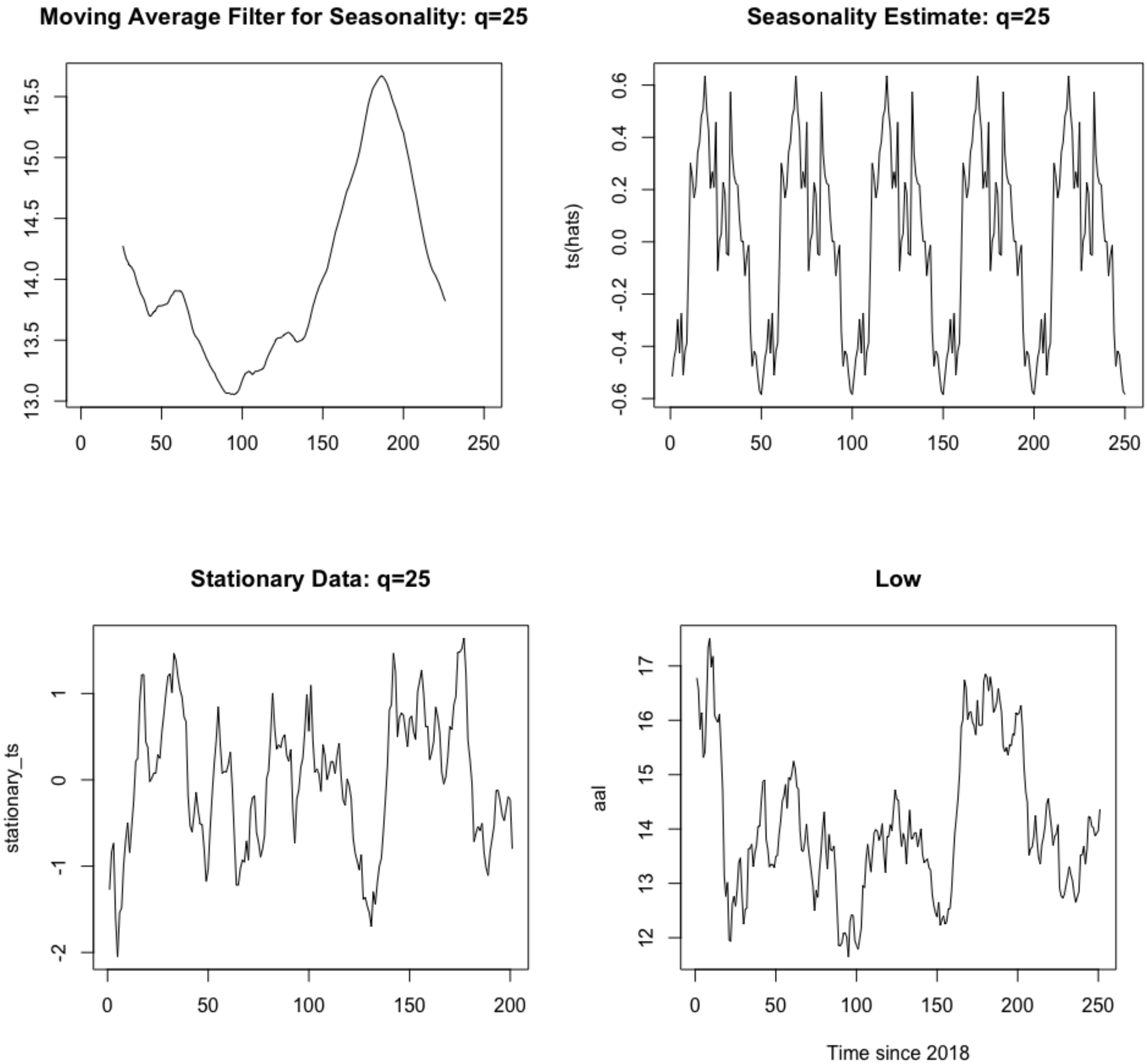

Figure 1: Top left: Applying a moving average filter of size q=25 to AAL time series. Top right: Seasonality estimate using the q=25 moving average filter. Bottom left: Stationary data with moving average filter and seasonality estimate removed. Bottom right: Initial time series for the Low in American Airlines stock prices over the past year

.

3.3 Analyzing Residuals

Stationary time series conditions: -Weakly stationary process: finite second moment, constant mean, ACF only dependent on h $\gamma(h)$. Our residuals seem to obey these properties.

3.3.1 Sample ACF

Cyclical behavior also shown here-reject white noise hypothesis. If white noise-ACF should be 1 at h=0 and around 0 for any other h. Does not say anything definitive about weak stationarity.

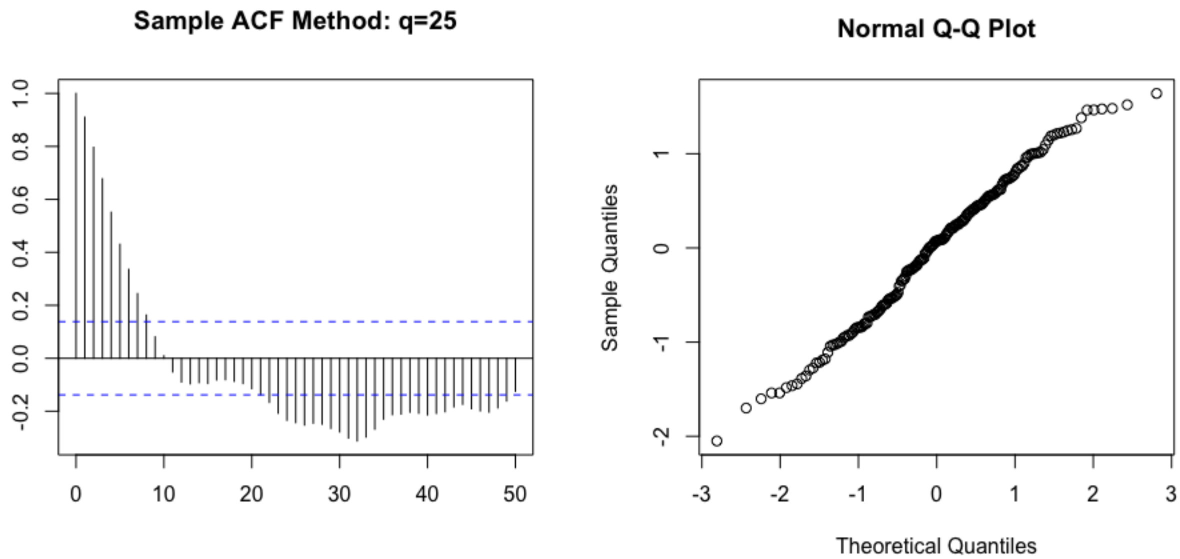

Figure 2: Left: Sample ACF for q=25 stationary data. Does not resemble white noise, but does seem to resemble weak stationarity. Right: qq plot for q=25: does not exactly resemble normality

.

3.3.2 Portmanteau Test

$\chi^2 = 556.21$ for this test and $p < 2.2e - 16$ with df=20. For this test: H_0: data is IID and H_a: data is not IID. We therefore reject H_0: data is not IID at a significance of $\alpha = 0.05$.

3.3.3 QQ-plot test and R-squared

Similar to the Portmanteau Test: H_0: data is Gaussian IID and H_a: data is not Gaussian IID. $R^2 = 0.9904714$ for this data: since P($R^2 < 0.987$) $= 0.05$, at a 0.05 significance level, we cannot reject Gaussian IID.

3.4 Fitting ARMA Models

ARMA is a way of describing a variety of weak stationary residuals. The ARMA(p,q) model is as follows [6]:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_q Z_{t-q} \tag{4}$$

Where ϕ and θ are the autoregressive and moving average weights, represented by the autoregressive and moving average polynomials:

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$$

$$\theta(z) = 1 + \theta_1 z - ... + \theta_q z^q$$

To fit an ARMA model, we first need to examine the autocorrelation function (ACF) and partial autocorrelation function (PACF). Based on the following properties, we can decide if the residuals follow an AR(p), MA(q) or ARMA(p,q) process [6]:

1. AR(p): ACF tails off, PACF cuts off after lag p
2. MA(q) ACF cuts off after lag q, PACF tails off
3. ARMA(p,q) both ACF and PACF tail off

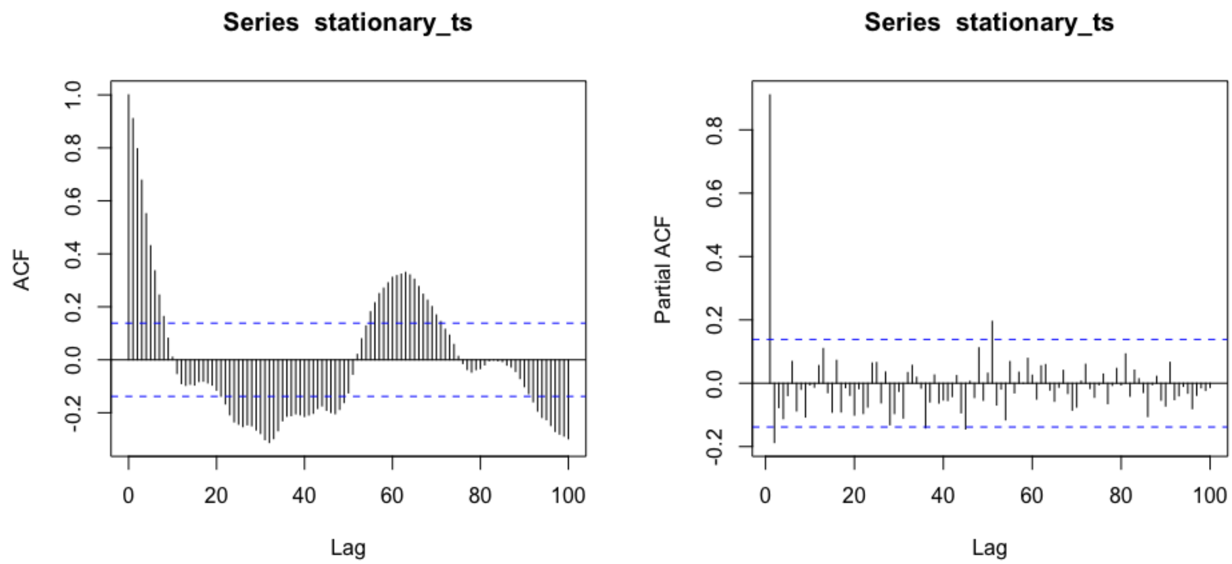

Figure 3: Left: ACF for q=25 residuals at lag=100. Right: pacf for q=25 residuals at lag=100. Both these processes seem to resemble an AR(1) or AR(2) process.

.

Since this could be an AR(1) or AR(2) process, we found the weights for both (with standard error)

3.4.1 AR(1) Process Fit

By fitting an AR(1) process to the residuals, the obtained weights are: $\phi_1 = 0.9151$, intercept = 0.002463 ± 0.02151 with $\sigma^2 = 0.09255$. And so the total process would be:

$$X_t = \phi_1 X_{t-1} + Z_t$$

Simulated values are shown 4:

3.4.2 AR(2) Process Fit

By fitting an AR(2) process to the residuals, the obtained weights are: $\phi_1 = 1.1503$, $\phi_2 = -0.2502$, intercept = -0.0003499 ± 0.02088 with $\sigma^2 = 0.08669$. And so the total process would be:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

Simulated values are shown 5:

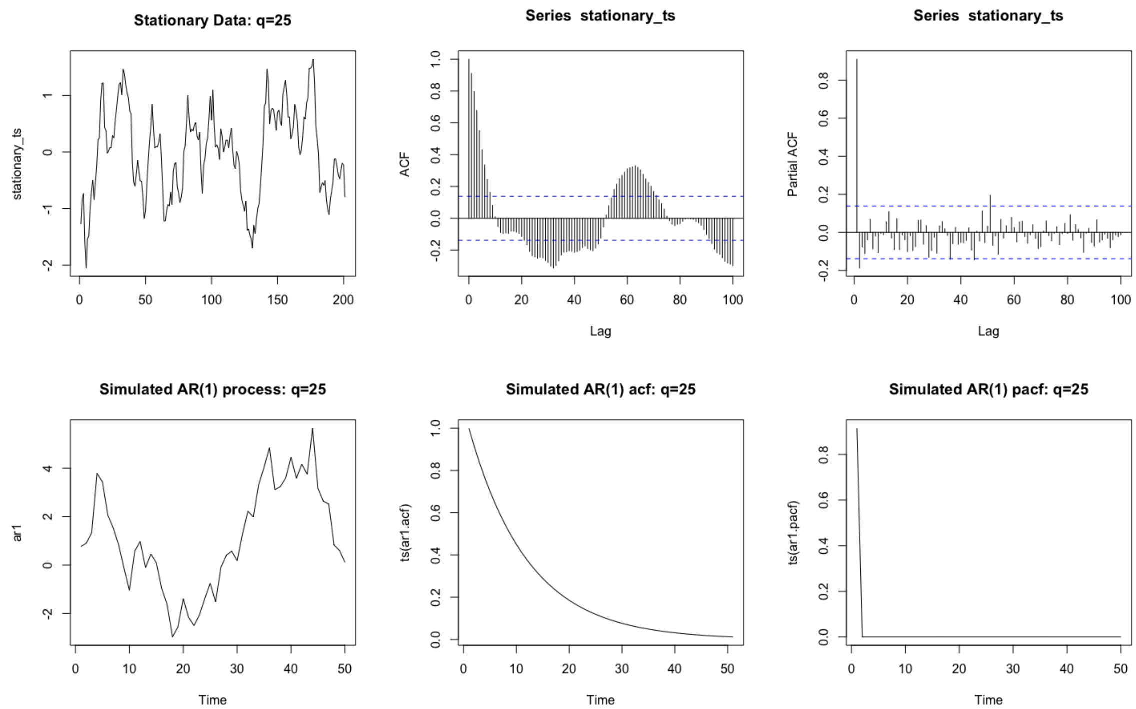

Figure 4: Top: Obtained residuals, ACF and PACF of q=25 estimation. Bottom: Simulation of AR(1) process. ACF and PACF don't exactly line up (especially the PACF which cuts off after lag 2

3.4.3 Parameter Estimation

For a sanity check on how the AR(1) and AR(2) weights ψ were found, we will estimate them via the Yule-Walker procedure and compare [6].

3.4.4 Observing Overall Residuals with Rough Component Removed

With both AR(1) and AR(2) processes estimated, we can take a look at the final residuals with the rough component removed and determine if it adheres to white noise 6:

It does not seem that the final residuals conform to white noise WN$(0, \sigma^2)$.

3.5 Predict Future Values

The overall prediction of the time series is the combined predictions of both the smooth and rough component: $\hat{m}_t$, $\hat{s}_t$ and X_t. Because we have estimated a pure AR processes, we will use the Durbin-Levinson algorithm for forecasting future values [6]. We will forecast over the next few days in May 2023:

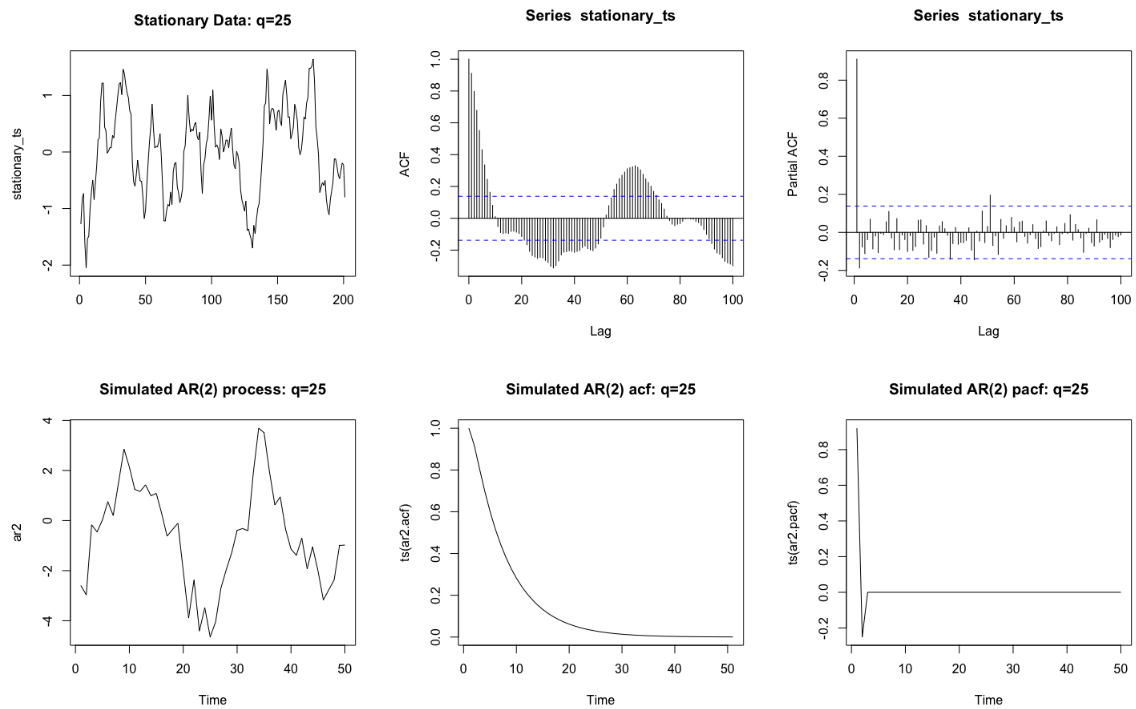

Figure 5: Top: Obtained residuals, ACF and PACF of q=25 estimation. Bottom: Simulation of AR(2) process. ACF and PACF don't exactly line up (especially the PACF which cuts off after lag 2

4 Discussion

The model is not perfect-by removing both smooth and rough component, we still did not get white noise. By looking at the PACF of the stationary data with the smooth component removed, there was no "clear" cutoff at lag=1,2. There were spikes roughly every 50 steps (roughly the period d). This leads us to think that our seasonality (and by implication, the pre-seasonal trend) estimates were incorrect. We used an AR(1) and AR(2) estimate, but perhaps the order p is much higher, or we estimated our smooth component incorrectly.

The deseasonalized data for the q=25 seasonality + moving average filter did not look much different than the initial data, yet the seasonality estimate was performed correctly. This leads me to think that: the moving average filter to estimate the pre-seasonality trend was inadequate. We redid the seasonality estimation with varying moving average filter sizes, but kept running into the same issue.

Another source of error that can be fixed is: a separate trend estimation should have been performed to the deseasonalized data, rather than use the same pre-seasonal trend estimate using the moving average filter of size q=25. We performed this calculation, re-estimating the trend component after the seasonality with a moving average filter of size q=10. The same error occurred, leading us to believe that modeling improvements to describe leftover dependence are necessary that may/may not be outside the scope of this course.

The trend function is not so clear-we postulated that it can be estimated by at least a 3rd degree polynomial due to the curvature.

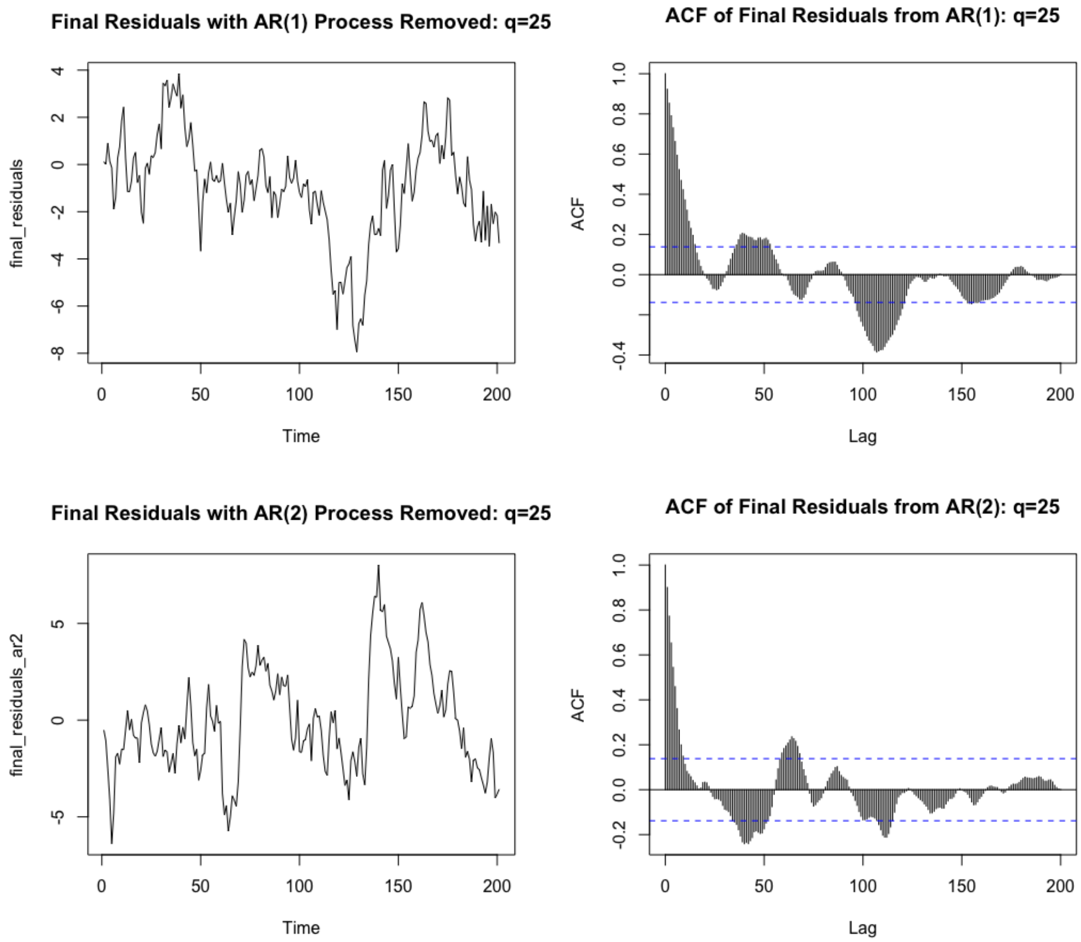

Figure 6: Top: Final residuals and ACF for AR(1) process. Bottom: Final residuals and ACF for AR(2) process. Neither of the ACFs conform to white noise

5 Conclusion

6 Spectral Analysis

Spectral analysis allows us to understand underlying periodicities, cyclic patterns, and dominant frequencies that can influence the price movements of the American Airlines stock.

To accurately conduct spectral analysis, we first use a moving average filter to detrend our data. After doing so, we remove all outliers in the detrended data.

We then apply one of the simplest forms of spectral analysis to the data, the periodogram. The periodogram identifies the dominant periods/frequencies in a time series that might explain the oscillation patterns in the observed data. This is completed by calculating the squared magnitude of the Fourier transforms by representing our time series as a sum of sine and cosine functions at the harmonic frequencies.

$$I(j/n) = |d(j/n)|^2 = \frac{1}{n}(\sum_{t=1}^{n} X_t cos(2\pi jt/n))^2 + \frac{1}{n}(\sum_{t=1}^{n} X_t sin(2\pi jt/n))^2 \tag{5}$$

References

[1] Yahoo Finance American Airlines Group Inc. (AAL), https://finance.yahoo.com

[2] The Complete Guide to Time Series Data, https://www.clarify.io/learn/time-series-data

[3] Taking stock of the pandemic's impact on global aviation, https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/taking-stock-of-the-pandemics-impact-on-global-aviation

[4] Time Series and Spectral Analysis, http://web.stanford.edu/class/earthsys214/notes

[5] 6.1 The Periodogram, https://online.stat.psu.edu/stat510/lesson/6/6.1

[6] STA 137 Lecture Notes, chapter 1-23, chapter 2-23, chapter 3-23