

# Impact of COVID-19 on Worldwide Aviation

Spencer Griffith, Philip Huang, Hao Jin, Qifeng Li

## Abstract

We analyze flights data relating to the COVID-19 pandemic using time series modeling techniques to explore how COVID-19 affected the airline industry. We make visualizations and analyze the impact of the virus on various aspects of air travel. We measure number of flights over time across different countries and see how various outbreaks affected flight numbers. We also visualize networks of the number of flights for busiest US airports and European airports measured on different time periods. Furthermore, we analyze metrics including number of passengers and revenue passenger miles. From examining the impact of COVID-19 on the airline industry from multiple angles, we see that the virus negatively affected flight numbers, passengers, revenue, and more. In order to see beyond what the data shows, we use three different models to forecast the future number of flights. The methods we use include Autoregressive Integrated Moving Average (ARIMA) models, Holt-Winters' Seasonal Method, and Vector Autoregressive (VAR) Models. The forecasts using ARIMA models and Holt-Winters' Method using `hw()` do not show any increase or decrease in number of flights. However, the `HoltWinters()` function and VAR models show a steady increase. The VAR model also shows a slight increase in the XAL Index.

**Keywords**— COVID-19, airline, flights, forecast, ARIMA, Holt-Winters, VAR

## I. INTRODUCTION

The COVID-19 pandemic had a profound impact on the aviation industry. Tens of thousands of flights have been canceled since the outbreak spread beyond China in January, 2020. Many countries implemented restrictions on air travel and some outright banned travel in and out of the country. Commercial air travel plummeted in 2020 and this affected airlines, airports, manufacturers, repair operators, and more. Economically, the airline industry lost billions and wiped out years of profit. Although 2021 and 2022 saw a revival in airline travel, the world is still vulnerable to different variants that could shut down the aviation industry yet again. In our project, we aim to illustrate the impact of COVID-19 on the airline industry and forecast how the future of flights will change with the ongoing pandemic.

To forecast the future number of flights, we use three different time series modeling techniques. The first is Autoregressive Integrated Moving Average (ARIMA) models, which are the most frequently used time series models. They perform very well for events that produce an immediate effect over a short period of time. The next technique we use is the Holt-Winters' Seasonal method, an exponential smoothing method which has been heavily used since 1960s. The last method are Vector Autoregressive (VAR) models. This kind of models is useful for predicting multiple time series variables using a single model. Using these three different techniques, we can forecast the number of flights and make conclusions based on the results.

### A. Data

1) **OpenSky flight data:** The most useful data source we use is the OpenSky flight data, which includes the origins and destinations of flights from many countries. We consider flight data related to the COVID-19 pandemic from Jan. 1 2019 to Dec. 31 of 2021 [33]. The OpenSky Network is a non-profit association based in Switzerland. The OpenSky Network consists of a multitude of sensors connected to the Internet by volunteers, industrial supporters, and academic/governmental organizations. The technology used by the OpenSky Network is the Automatic Dependent Surveillance-Broadcast (ADS-B). The ADS-B allows airplanes to determine their position and velocity using GPS. Airplanes then periodically broadcast this and other information on the 1090 MHz radio frequency. To obtain this valuable information at a large scale, OpenSky Network operates a network of ADS-B receivers around the globe and harvest the data via the Internet. OpenSky Network cannot provide every global flight movement in the dataset but only those ADS-B-equipped aircraft "seen" within its coverage. OpenSky has very good coverage in Europe and North America. Therefore, we focus on analyzing US and European airports and airlines.

The original data consisted of origin and destination in terms of ICAO codes (a four-letter code designating aerodromes around the world), so we first downloaded a table that had ICAO codes and their respective country names and mapped the corresponding country origin and destination into our table. After that, we subset the rows that does not have a missing value for country origin and country destination.

2) **BTS data:** Monthly number of passengers in all US airports [25], and monthly revenue passenger miles in all US air carrier domestic and international [26], collected by Bureau of Transportation Statistics (BTS).

3) **COVID-19 data:** Daily COVID-19 data across multiple countries [2], including the province/state, country/region, and confirmed cases counts, collected by the Center for Systems Science and Engineering at Johns Hopkins University. This comes from many different sources such as World Health Organization, ECDC, The Mercury News, US CDC, and more. This includes data from 2020 up to the current day in 2022.

4) **The NYSE Arca Airline Index (XAL):** The NYSE Arca Airline Index ("XAL" or "Index") is an equal-dollar weighted index designed to measure the performance of highly capitalized companies in the airline industry. The XAL Index tracks the price performance of selected stocks or ADRs of major U.S. and overseas airlines. The XAL Index was established with a benchmark value of 200.00 on October 21, 1994. The historical data we use in this project can be retrieved from yahoo finance [10].

\*This work was not supported by any organization

## B. Time Series Preliminaries and Notation

A sequence of random variables  $\{Y_t, t = 0, \pm 1, \pm 2, \dots\}$  is called a stochastic process and is served as a model for a set of observed time series data. It is convenient to refer to  $\{Y_t\}$  itself as a time series. A time series  $\{Y_t\}$  is said to be *weakly stationary* (or *second-order stationary* or *covariance stationary*) if  $\mathbb{E}(Y_t^2) < \infty$  and, for any integer  $k$ , neither  $\mathbb{E}(Y_t)$  nor  $\text{cov}(Y_t, Y_{t+k})$  depend on  $t$ . For simplicity, we call  $\{Y_t\}$  *stationary* if it is weakly stationary, i.e.  $\{Y_t\}$  has finite and time-invariant first two moments. Sometimes, we can transform a *nonstationary* series to make it stationary. We define the *order of integration* of a series as the minimum number of times it must be differenced to make it stationary. Hence stationary series are said to be integrated of order zero,  $I(0)$ . Similarly, an  $I(1)$  nonstationary series can be made stationary by taking the first difference.

For a weakly stationary time series  $\{Y_t\}$  with common mean  $\mu = \mathbb{E}(Y_t)$ , the *autocovariance function* (ACVF) is defined as

$$\gamma(k) = \text{cov}(Y_t, Y_{t+k}) = \mathbb{E}\{(Y_t - \mu)(Y_{t+k} - \mu)\}, \quad (1)$$

and the *autocorrelation function* (ACF) is defined as

$$\rho(k) = \text{corr}(Y_t, Y_{t+k}) = \frac{\gamma(k)}{\gamma(0)} \quad (2)$$

for  $k = 0, \pm 1, \pm 2, \dots$ . Note that  $\gamma(0) = \text{var}(Y_t)$  is independent of  $t$ . In practice, we use an observed sample  $Y_1, \dots, Y_T$  to estimate ACVF and ACF by the *sample ACVF* and *sample ACF*. They are defined as follows

$$\hat{\gamma}(k) = \frac{1}{T} \sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}), \quad \hat{\rho} = \hat{\gamma}(k)/\hat{\gamma}(0), \quad (3)$$

where  $\bar{Y} = T^{-1} \sum_{1 \leq t \leq T} Y_t$ . It is often useful to test whether a sample autocorrelation is significantly different from zero. That is, a test of  $H_0 : \rho(k) = 0$  against the alternative  $H_1 : \rho(k) \neq 0$ . Tests of this nature are useful for constructing models and for checking whether the errors in an equation might be autocorrelated or serially correlated. The test statistic for this test is relatively simple. When the null hypothesis is true, sample autocorrelation  $\hat{\rho}(k)$  has an approximate normal distribution with mean zero and variance  $1/T$ . Thus, a suitable test statistic is

$$Z = \frac{\hat{\rho}(k) - 0}{\sqrt{1/T}} = \sqrt{T}\hat{\rho}(k) \stackrel{a}{\sim} N(0, 1). \quad (4)$$

At a 5% significance level, we reject  $H_0 : \rho(k) = 0$  when  $\sqrt{T}\hat{\rho}(k) \geq 1.96$  or  $\sqrt{T}\hat{\rho}(k) \leq -1.96$ .

The ACF  $\rho(k)$  measures the correlation between  $Y_t$  and  $Y_{t+k}$  regardless of their relationship with the intermediate variables  $Y_{t+1}, \dots, Y_{t+k-1}$ . The *partial autocorrelation function* (PACF) at lag  $k$ , denoted by  $\pi(k)$ , is the conditional correlation between  $Y_t$  and  $Y_{t+k}$  given the intermediate variables  $Y_{t+1}, \dots, Y_{t+k-1}$ . The precise definition is somewhat technical. Here we present a way for estimating PACF from data. For each given  $k$ , run the linear regression by minimizing

$$\mathbb{E}(Y_{t+k} - \beta_0 - \beta_1 Y_{t+k-1} - \cdots - \beta_k Y_t)^2$$

with respect to  $\beta$ . The regression coefficient depend on  $k$  and are denoted by  $b_{k0}, \dots, b_{kk}$ . The PACF is then  $\pi(k) = b_{kk}$ , the last regression coefficient in the fit.

A very specific class of processes that plays a similar role to zero in the number theory is the *white noise*. When  $\rho(k) = 0$  for any  $k \neq 0$ ,  $\{\varepsilon_t\}$  is called a white noise and is denoted by  $\varepsilon_t \sim WN(\mu, \sigma^2)$ , where  $\sigma^2 = \gamma(0) = \text{var}(\varepsilon_t)$ . In other words, a white noise is a sequence of random variables with (1) same mean, (2) same variance, and (3) no serial correlation. A simple and frequently used test for white noise is *Ljung-Box portmanteau test*. The Ljung-Box  $Q_m$ -statistic seeks to test the null joint hypothesis  $H_0 : \rho(1) = \cdots = \rho(m) = 0$ . Rejecting this hypothesis means that some or all autocorrelations up to order  $m$  are different from zero. The Ljung-Box Q-statistic, has the following formula

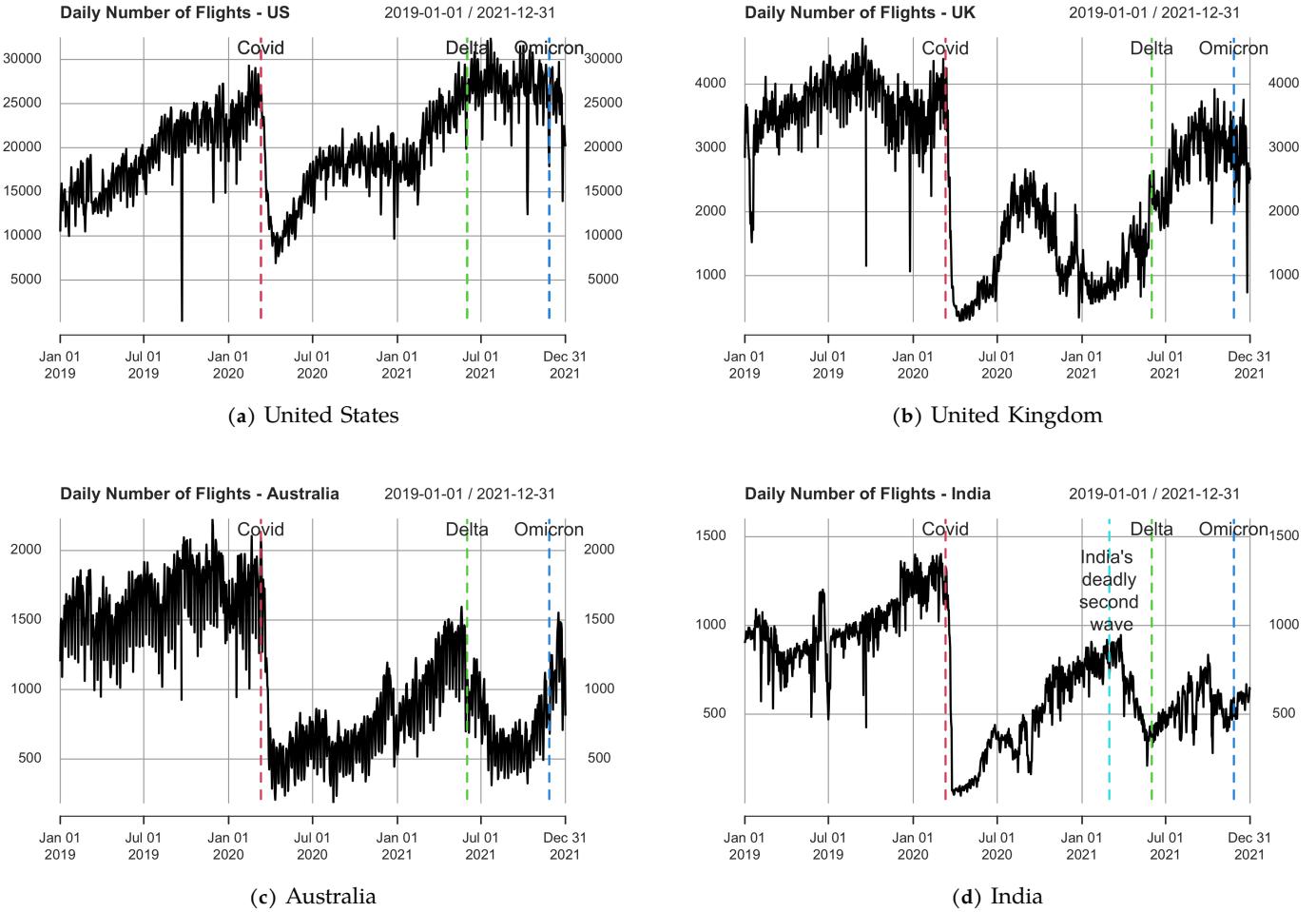
$$Q_m = T(T+2) \sum_{j=1}^m \frac{1}{T-j} \hat{\rho}(j) \longrightarrow \chi_m^2 \quad (5)$$

and, under the null, its probability density function is a chi-square with  $m$  degrees of freedom. For significance level  $\alpha$ , the critical region for rejection of the hypothesis of randomness is:  $Q_m > \chi_{1-\alpha, m}^2$  where where  $\chi_{1-\alpha, m}^2$  is the  $(1 - \alpha)$ -quantile of the chi-squared distribution with  $m$  degrees of freedom. The above test is sensitive to the choice of  $m$  [9]. In practice, we often carry out this test with different values of  $m$ .

The vector counterpart of this preliminary note can be found in Appendix A.

## C. Organization

Through the analysis of the airline industry and COVID-19, this paper explores the impact of the virus and looks into where the future might lead for air travel. Section II investigates the impact of COVID-19 on various aspects of the airline industry. We analyze the number of flights across different countries, number of passengers in the US domestically and internationally, US revenue passenger miles, flights for the ten busiest US airports and their European counterparts, and compare flights and confirmed cases in the US and UK. In Section III, we lay out the methods we use for forecasting the number of flights. The methods include Autoregressive Integrated Moving Average (ARIMA) models, Holt-Winters' Seasonal Method, and Vector Autoregressive (VAR) Models. Each method has their strengths and weaknesses, and produces slightly different results accordingly. This section explains what exactly each technique is, how to estimate parameters, model selections and diagnostics, and how to make forecasts. The next segment, section IV, shows the forecasting results from these forecasting methods along with some model diagnostics. Section V is our conclusion, in which we reflect on our analysis and communicate the results of our project.



**Figure 1:** Daily number of flights (domestic and international) in different countries from 2019-2021 with line breaks for the COVID (SARS-CoV-2), the Delta variant, and the Omicron variant outbreaks (dates cited from CDC).

## II. THE IMPACT OF COVID-19

### A. Number of Flights

One way the impact of COVID can be measured is the time it takes until the number of flights after the initial outbreak reached the point it was before the outbreak. From the graphs in Figure 1, the US has actually risen beyond where it was before the outbreak. Compare this to the UK, which has not yet reached the numbers of flights it was at before COVID. Australia and India also have not recovered since COVID has begun.

On each of the graphs, there are line breaks indicating when different strains of the virus occurred. For every country, the initial COVID outbreak halted nearly all flight travel. After that, we included the point at which the Delta and Omicron variants occurred. These points have a different effect for each country, where countries like the US, UK, and India were not affected that much. However, for a country like Australia, it seems like the Delta variant had a significant effect on flight numbers. India is slightly unique, as the country had its own second deadly wave of COVID-19, and flight numbers were hit heavily by this.

The US is very surprising in that it looks like COVID hardly slowed down flights at all and perhaps even contributed to raising the flight numbers towards the end of 2021. This could be because the COVID policies in the US took a while to be enacted and were not enforced as strictly as other places. It seems like all of the other countries have a similar graph where after the initial drop the flight numbers increased a little bit but generally people and the government were still wary of the virus and perhaps did not open airports fully.

### B. Change Detection

In time series analysis, change point detection tries to identify times when the probability distribution of a stochastic process or time series changes. In general, the problem concerns both detecting whether or not a change has occurred or whether several changes might have occurred, and identifying the times of any such changes.

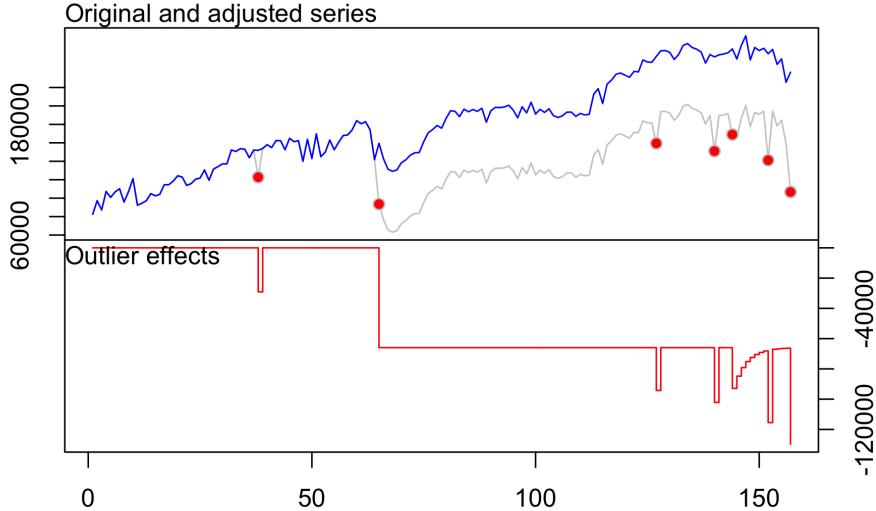
Here we use time series outlier detection to detect changes in time series. Chen and Liu's [5] procedures are popular time series outlier detection methods. The `tsoutlier` package in R uses Chen and Liu's method to detect outliers. Five types of outliers can be detected: Additive Outlier (AO), Innovation Outlier (IO), Level Shift (LS), Temporary change (TC), and Seasonal Level Shift (SLS).

For the time series of weekly number of flights in the US, `tsoultlier::tso` function identifies seven outliers; see table I. Here `ind` or `time` represent week number, where the week 1 is defined as the first week of 2019. The package also provides

	type	ind	time	coefhat	tstat
1	AO	38	38.00	-29094.79	-4.00
2	LS	65	65.00	-65960.87	-7.74
3	AO	127	127.00	-28305.50	-3.89
4	AO	140	140.00	-36211.77	-4.97
5	TC	144	144.00	-26908.69	-3.34
6	AO	152	152.00	-47968.14	-6.53
7	AO	157	157.00	-63585.11	-6.94

**Table I:** Change points in weekly number of flights in the US.

nice plots. Figure 2 shows where the outliers are (red dots) and also what would have happened if there were no outliers (the blue curve). The most noteworthy outlier is the second one in table I, which is of a *LS* type. This changing point of the time series happened in week 65, which corresponds to mid-March 2020 i.e. the initial outbreak of the COVID.



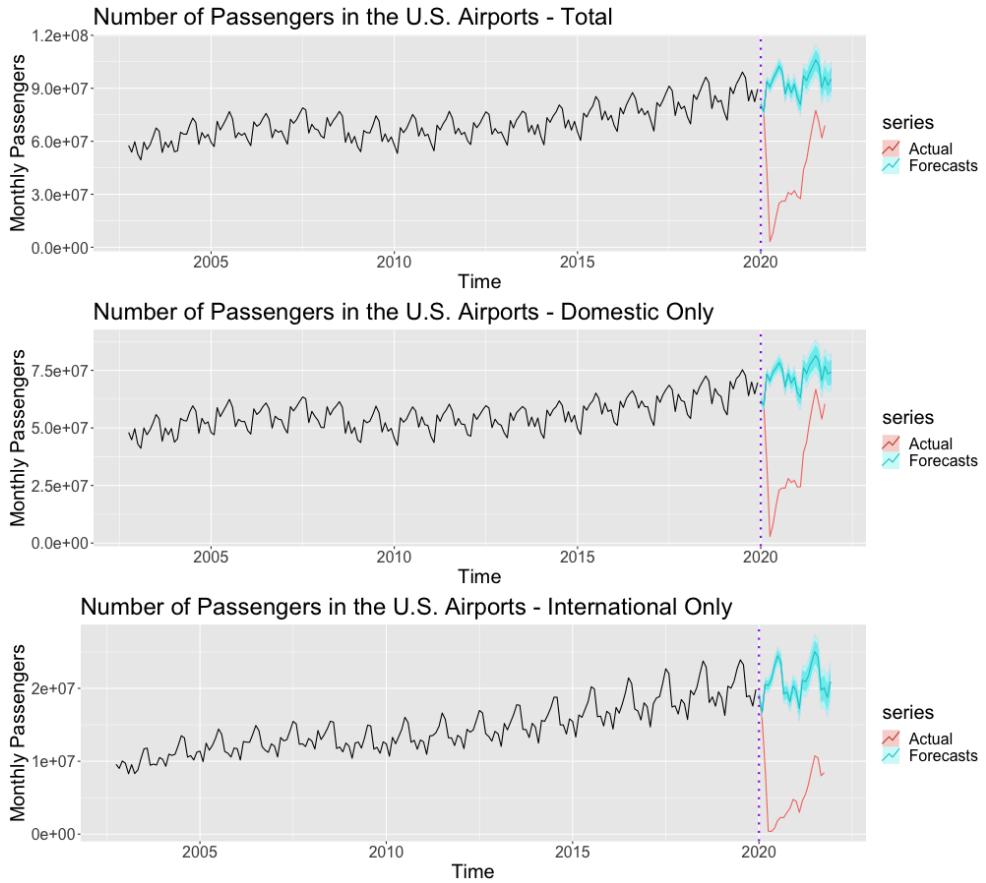
**Figure 2:** Change points in weekly number of flights in the US.

We have also used R function `strucchange::breakpoints` to detect level shifts. The result shows that there was some level shift happening around week 64 to 65 (mid-March 2020), which is consistent with the result given by Chen and Liu's method.

### C. Number of Passengers

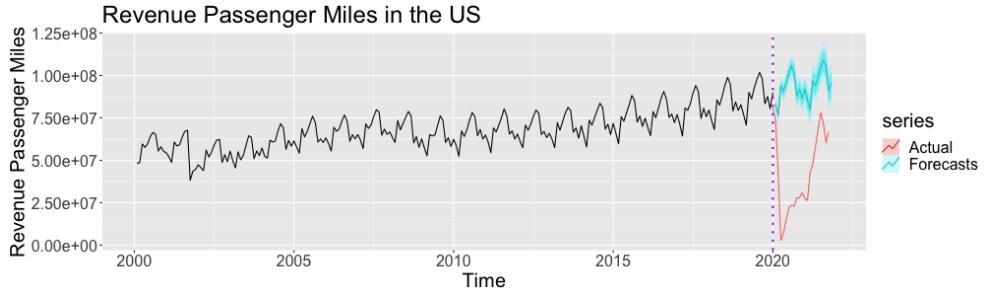
We fit a seasonal ARIMA model ( $\text{ARIMA}(0,1,2)(2,1,2)_{12}$ ), which we will discuss in detail later, on the monthly passenger data from October 2002 to December 2019; see Figure 3 (The dark shaded region shows 80% prediction intervals. That is, each future value is expected to lie in the dark shaded region with a probability of 80%. The light shaded region shows 95% prediction intervals. These prediction intervals are a useful way of displaying the uncertainty in forecasts. In this case the forecasts are expected to be accurate, and hence the prediction intervals are quite narrow.) We then compare the forecasted values to the real monthly passenger data in 2020 and 2021. We can clearly see from the plot that the number of passengers almost quartered in 2020 and halved in 2021. However, it is interesting to observe that COVID did not wipe out seasonal patterns from the time series.

Similarly, we split the number of passengers in US airports into Domestic and International flights separately. By fitting an  $\text{ARIMA}(0,1,2)(2,1,2)[12]$  to the domestic series and an  $\text{ARIMA}(0,1,1)(2,1,0)[12]$  to the international series, we also compare forecasted values to real data from 2020 and 2021 as before. We can see that domestic flights held much more passengers than international flights. Also, it is interesting to note that by the end of 2021, domestic flights almost reached the predicted number of passengers while international flights were nowhere near the forecasted values. Domestic flights were most likely not affected as much because the US did not have as strict COVID policies as other countries. If people wanted to travel to other states, there were very little obstacles in their way. On the other hand, many international countries allowed limited air travel into and from their country, making it hard for international travel. Although it is hard to tell, there are still seasonal patterns in both domestic and international passenger plots.



**Figure 3:** Number of air passengers in the US airports, with ARIMA forecasts and actual data from 2020 onward.

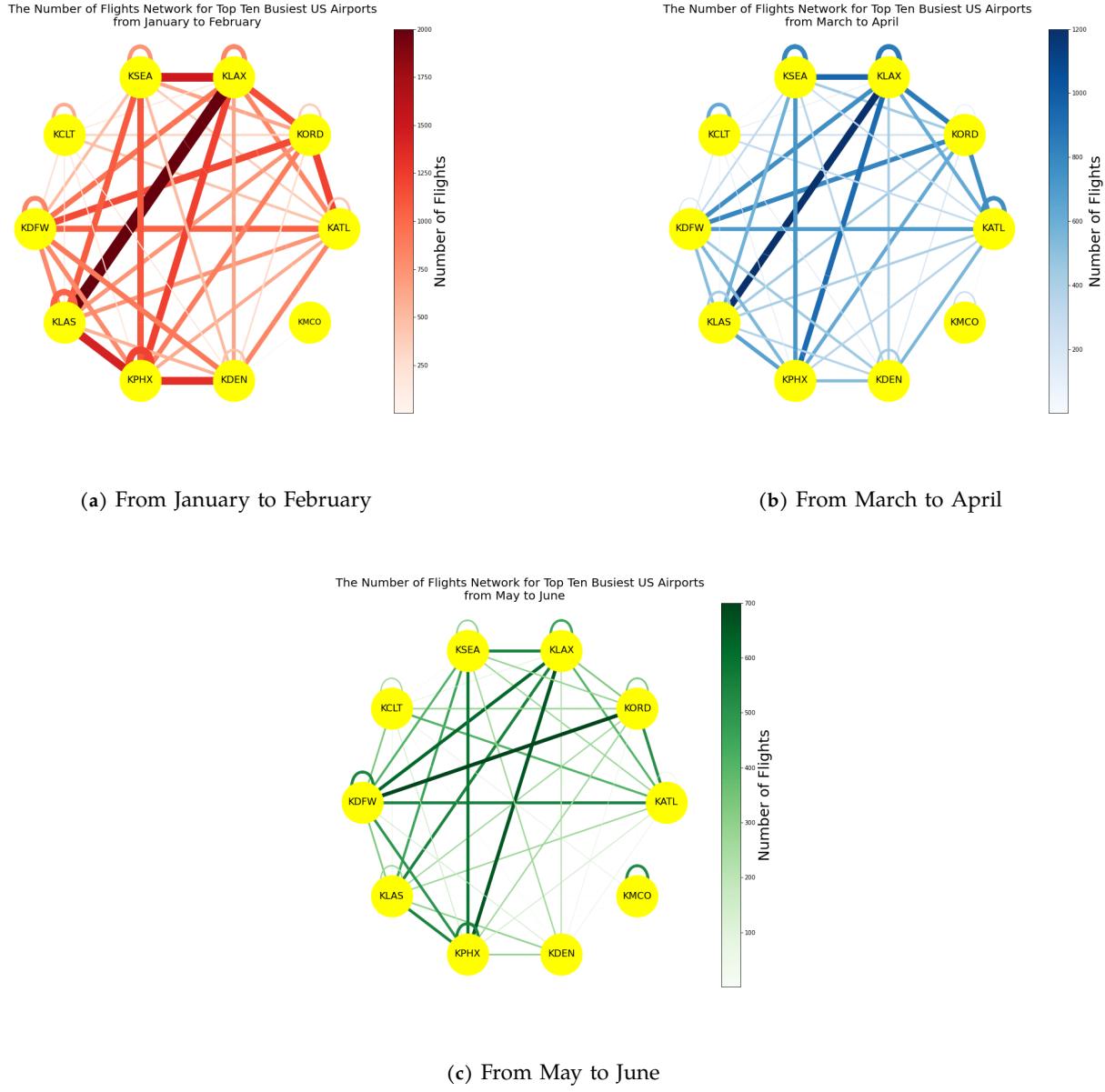
#### D. Revenue Passenger Miles



**Figure 4:** Revenue passenger-miles (RPM) in the US, with ARIMA forecasts and actual data from 2020 onward.

Revenue passenger-miles (RPM) are a measure of the volume of air passenger transportation. A revenue passenger-mile is equal to one paying passenger carried one mile. In the graph from Figure 4, we fit an ARIMA model to the revenue passenger miles in the US, including domestic and international travel. We also compare forecasted values to real data from 2020 and 2021 as in other graphs. It is clear that there is a seasonal pattern not unlike other data, and this continues even after the COVID outbreak. The beginning of 2020 saw a drop to near zero revenue passenger miles. Towards the end of 2021, it is interesting to note that the actual revenue passenger miles almost reached the forecasted values. So while COVID initially had a devastating effect on RPM, it seems that the US has almost bounced back from the damaging consequences of the virus.

### E. Network Visualization

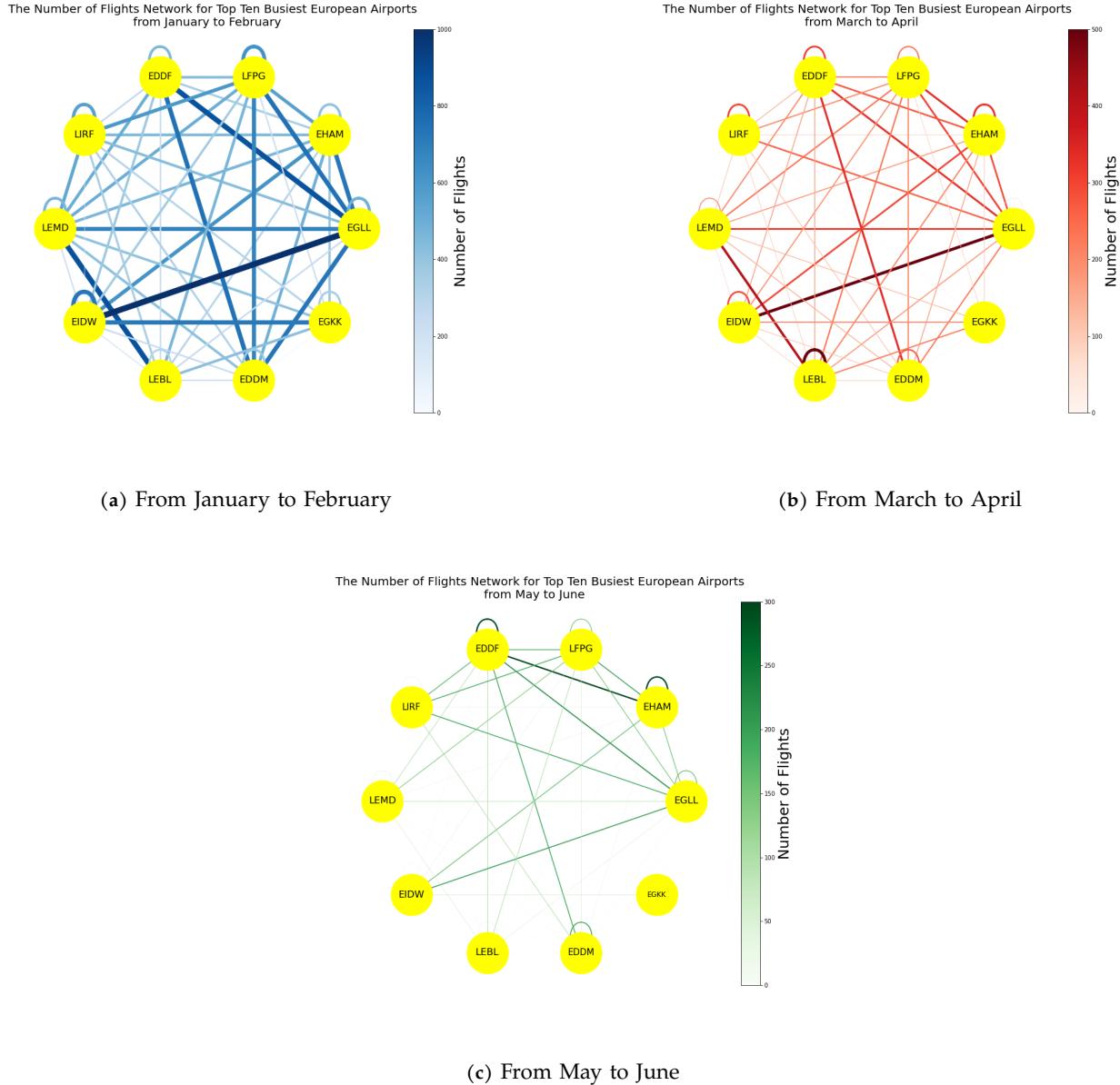


**Figure 5:** The number of flights network for top ten busiest US airports before and after the outbreak of COVID-19.

We looked at the top ten busiest domestic airports in the United States according to Wikipedia and then visualized the number of flights between the ten airports during the months of January and February, March and April, and May and June of 2020. The names of these airports along with their ICAO codes can be found in Table II. The circular graphs between these two-month intervals are consistent in that the width represents the number of flights from each airport to another, with wider edge width meaning more flights between the two airports. The color bar represents the range of number of flights from the minimum to the maximum number of flights that occurred during that two-month period. The minimum and maximum number of flights varied between different two-month periods, so we decided to use different colors to prevent confusion in comparing the colors between the graphs. From January to February 2020, the edge width of many of the connections are pretty thick, meaning domestic flights are still occurring rather frequently. Late December to early January was when news first broke out that coronavirus had appeared in Wuhan, China. However, from the networks, it does not seem like domestic flights have been dramatically reduced. Then for March to April, we can see that the number of flights start to decrease as the edge widths became smaller. This was when coronavirus started to hit the US hard and cases were increasing. Then on May to June, we can see a thinning of the edges, which probably signifies the policy changes as politicians were starting to take coronavirus more seriously.

At the start of the pandemic, there seems to have been a sentiment in the United States that because the virus originated from China that it wouldn't have a huge effect on the United States. The emphasis on a strong economy likely prevented many domestic flights from being canceled early on. People continued to work and this could have contributed to the spread of the

virus, as there does not seem to be a drastic reduction in the number of domestic flights from January to February. What this potentially shows is that in an ever-more connected world we live in, actions and policy changes need to be taken early on in order to aid in preventing the spread of coronavirus. In this case, there should have been a drastic reduction in the number of flights domestically, not just internationally. In the future when another pandemic occurs, hopefully we will be more prepared and shut down flights early on as the graph seems to indicate a delayed response in reducing the number of flights domestically.



**Figure 6:** The number of flights network for top ten busiest European airports before and after the outbreak of COVID-19.

Then we looked at the top ten busiest European airports according to Wikipedia. Like the United States, there seems to have been a reduction in flights as the two-month periods progressed. Europe seems to have taken action similar to the United States, and there also seems to have been a delayed response in flight cancellation from January to February compared to March and April.

ICAO Code and Corresponding Airports			
ICAO code for 10 US airports		ICAO code for 10 European airports	
KSEA	Seattle-Tacoma Int Airport	EDDF	Frankfurt Airport (Germany)
KLAX	Los Angeles Int Airport	LFPG	Paris Charles de Gaulle Airport (France)
KORD	O'Hare Int Airport	EHAM	Amsterdam Airport Schiphol (Netherlands)
KATL	Hartsfield-Jackson Atlanta Int Airport	EGLL	Heathrow Airport (UK)
KMCO	Orlando Int Airport	EGKK	Gatwick Airport (UK)
KDEN	Denver Int Airport	EDDM	Munich Int Airport (Germany)
KPHX	Phoenix Sky Harbor Int Airport	LEBL	Josep Tarradellas Barcelona-El Prat Airport (Spain)
KLAS	Harry Reid Int Airport	EIDW	Dublin Airport (Ireland)
KDFW	Dallas/Fort Worth Int Airport	LEMD	Adolfo Suárez Madrid-Barajas Airport (Spain)
KCLT	Charlotte Douglas Int Airport	LIRF	Leonardo da Vinci Int Airport (Italy)

Table II: ICAO code and corresponding airports.

#### F. Number of Confirmed Cases

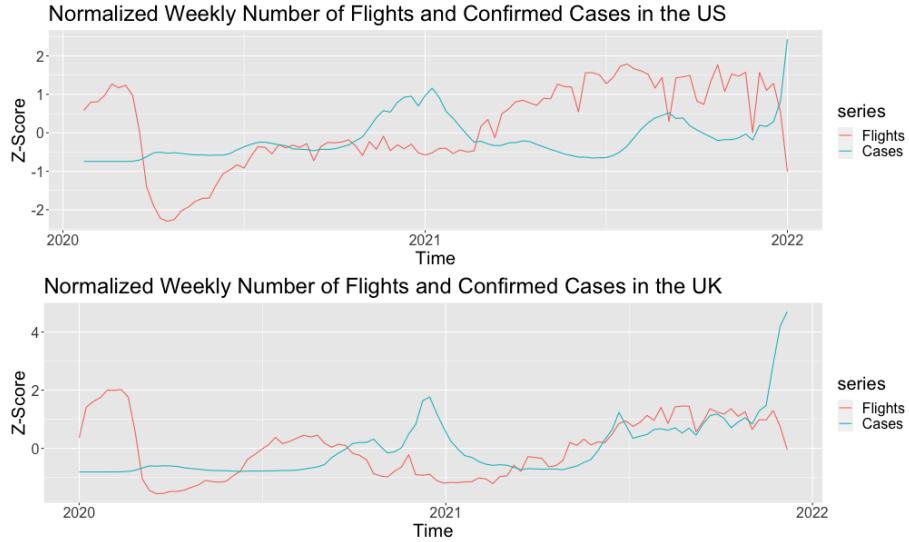


Figure 7: Normalized weekly number of flights and weekly number of confirm cases in the US and the UK.

Based on the graphs in Figure 7, it seems like there is a “negative correlation” between number of COVID cases and number of flights. This means that when COVID cases are increasing, the number of flights goes down. This is to be expected as there are government regulations and COVID scares that prevent people from flying. Both the US and UK have similar graphs, although the UK’s confirmed cases stretches to almost 5 standard deviations away, while the US stays within 2 standard deviations. It looks like the UK recovered slightly faster as the number of flights reached above a z-score of 0 faster than the US in mid 2020. However, when the second big increase in cases happened, the US seemed to recover faster in early 2021. The end of 2021 shows a very similar picture for both countries as the final spike in confirmed cases occurs.

### III. FORECASTING NUMBER OF FLIGHTS - METHODS

#### A. Autoregressive Integrated Moving Average (ARIMA) Models

1) **Stationary autoregressive moving average (ARMA) models:** One of the most frequently used time series models is the stationary autoregressive moving average (ARMA) model. Combining AR and MA together (key properties of AR and MA models are derived and summarized in Appendix B.), a general *autoregressive moving average* (ARMA) model with the order  $(p, q)$  has the form

$$Y_t = c + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad (6)$$

where  $\varepsilon_t \sim WN(0, \sigma^2)$ , and  $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  are parameter. Let

$$\Theta(x) = 1 + \theta_1 x + \cdots + \theta_q x^q \quad \text{and} \quad \Phi(x) = 1 - \phi_1 x - \cdots - \phi_p x^p.$$

Then, model (6) can be more compactly written as

$$\Phi(L)Y_t = \Theta(L)\varepsilon_t.$$

To ensure that  $(p, q)$  are the genuine orders of the model, we assume that the two equations  $\Theta(x) = 0$  and  $\Phi(x) = 0$  do not have any common roots. Otherwise, we can cancel the common factors from the above equation. Based on the discussions on MA and AR, the following assertions hold [9].

**Stationarity** When the  $p$  roots of the AR characteristic equation  $\Phi(x) = 0$  are outside of the unit circle, equation (6) defines a stationary process which admits an  $\text{MA}(\infty)$  representation

$$Y_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon \sim \text{MA}(\infty)$$

i.e. such a stationary solution is also causal. Furthermore  $\mathbb{E}(Y_t) = c/(1 - \phi_1 - \dots - \phi_p)$ , and the ACF  $\rho(k) \rightarrow 0$  at an exponential rate as  $k \rightarrow \infty$ .

**Invertibility** When the  $q$  roots of the MA characteristic equation  $\Theta(x) = 0$  are outside of the unit circle, equation (6) defines a invertible process in the sense that  $Y_t \sim \text{AR}(\infty)$ .

When stationarity and invertibility hold, we have three very important properties of ARMA models: (1) they are mean-reverting processes so that the data may fluctuate but, sooner or later, the process reverts to its unconditional mean, (2) they are short memory processes so that information in the far and very far past is irrelevant to produce a forecast today, and (3) their multi-step forecast converges to the unconditional mean of the process when the forecast horizon is far into the future [12]. We may understand (3) as a consequence of (1) and (2).

For a nonstationary time series, one often takes the difference and hopes the resulting time series is stationary. *Autoregressive integrated moving average* (ARIMA) models from a useful class of nonstationary models. We say  $Y_t \sim \text{ARIMA}(p, d, q)$  if  $\nabla^d Y_t$  is a stationary ARMA( $p, q$ ), where  $\nabla$  denotes the difference operator, and for any integer  $d \geq 2$ ,  $\nabla^d Y_t \equiv \nabla(\nabla^{d-1} Y_t)$ . Notice that an ARIMA( $p, 0, q$ ) model is the same as an ARMA( $p, q$ ) model. ARIMA( $p, 0, 0$ ), ARMA( $p, 0$ ), and AR( $p$ ) models are the same. Similarly, ARIMA( $0, 0, q$ ), ARMA( $0, q$ ), and MA( $q$ ) models are the same.

2) **Fitting ARMA models:** When the distribution of  $\varepsilon_t$  is known, the maximum likelihood estimation is more efficient than the least squares estimation [9]. The most common assumption in time series analysis is that  $\varepsilon$  is independent and  $N(0, \sigma^2)$  in model (6). For simplicity, we assume  $c = 0$  in (6), i.e.  $\mu = \mathbb{E}(Y_t) = 0$ . Under these assumptions  $\mathbf{Y}_T \equiv (Y_1, \dots, Y_T)$  are jointly normal with common mean 0 and covariance matrix  $\Sigma$ , whose  $(i, j)$ -th element is  $\gamma(i-j)$  that depends on the model parameters  $(\theta, \phi, \sigma^2)$  and is written as  $\Sigma(\theta, \phi, \sigma^2)$ .

To derive the maximum likelihood estimator, we need to find the density of  $\mathbf{Y}_T$ . From the multivariate analysis [3], the density of  $\mathbf{Y}_T \sim N(\mu, \Sigma)$  is given by

$$f(\mathbf{Y}_T | \mu, \Sigma) = \frac{1}{(2\pi)^{T/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{Y}_T - \mu)^\top \Sigma^{-1} (\mathbf{Y}_T - \mu) \right\}. \quad (7)$$

Regarding  $f(\mathbf{Y}_T | \mu, \Sigma)$  as a function of parameters  $\mu$  and  $\Sigma$  results in the likelihood function with the log-likelihood function  $\ell(\mu, \Sigma)$ . Dropping the constant factors in (7),

$$-2\ell(\mu, \Sigma) = \ln |\Sigma| + (\mathbf{Y}_T - \mu)^\top \Sigma^{-1} (\mathbf{Y}_T - \mu).$$

In our application,  $\mu = 0$  and  $\Sigma = \Sigma(\theta, \phi, \sigma^2)$  so that the unknown parameters are  $\theta, \phi$ , and  $\sigma^2$ . Thus, the maximum likelihood estimator (MLE) for  $(\theta, \phi, \sigma^2)$  is then defined as

$$(\hat{\theta}, \hat{\phi}, \hat{\sigma}^2) = \arg \min_{\theta, \phi, \sigma^2} \{ \mathbf{Y}_T^\top \{ \Sigma(\theta, \phi, \sigma^2) \}^{-1} \mathbf{Y}_T + \ln |\Sigma(\theta, \phi, \sigma^2)| \}. \quad (8)$$

The Gaussian MLE  $(\hat{\theta}, \hat{\phi}, \hat{\sigma}^2)$  defined in (8) is asymptotically unbiased, normal with the limit covariance matrix determined by the ACF of  $Y_t$ , provided that  $Y_t$  is causal and invertible, while  $\varepsilon_t$  is not necessarily normal [8].

3) **Stationarity:** The slow decay of the sample ACF is often taken as a symptom of nonstationary time series due to the existence of either dynamic or deterministic trend. Differencing is often adopted, which eventually leads to the fitting of the data with an appropriate ARIMA model. However, we would also like to apply some formal statistical tests to assess if differencing is necessary. There are many tests for determining whether a series is stationary or nonstationary. The *augmented Dickey-Fuller* (ADF) test is one of such tests. Details of this test can be found in Appendix C.

4) **Seasonality:** Some time series exhibit periodic behavior. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models. It is written as  $\text{ARIMA}(p, d, q)(P, D, Q)_m$ , where  $m = \text{number of observations per year}$ . We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model. The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve lags of the seasonal period. For example, an ARIMA(1, 1, 1)(1, 1, 1)<sub>4</sub> model (without a constant) is for quarterly data ( $m = 4$ ), and can be written as [23]

$$(1 - \phi L)(1 - \Phi_4 L^4)(1 - L)(1 - L^4)Y_t = (1 + \theta_1 L)(1 + \Theta_4 L^4)\varepsilon_t. \quad (9)$$

5) **Model diagnostic:** The formulation of ARMA models is a statistical assumption for unknown data generating processes. Whether it is appropriate for a given time series requires a validation. Model diagnostic checks the goodness-of-fit of a model with the data. The key idea is simple: if ARMA model (6) is adequate for a given data set, the residuals

$$\hat{\varepsilon} = Y_t - \sum_{j=1}^p \hat{\phi}_j Y_{t-j} - \sum_{i=1}^q \hat{\theta}_i \hat{\varepsilon}_{t-i}, \quad t = p+1, \dots, T, \quad (10)$$

should behave like white noise. We may calculate the above recursion with the initial values  $\hat{\varepsilon}_{p+1-i} = 0$  for  $i = 1, \dots, q$ .

An intuitive method for residual analysis is visual diagnostics. It is useful to plot the residuals  $\hat{\varepsilon}_t$  against time  $t$ . Such a plot allows us to examine whether there is any time trend or strong serial correlation in the residuals. If our ARMA model is adequate, the residuals should behave like patternless white noise. We often plot the standardized residuals, i.e. the residuals divided by their standard errors. If  $\hat{\varepsilon}_t$  are normal, then approximately 95% of points in a standardized residual plot should be within the range  $[-2, 2]$ . We may also look into the ACF and PACF plots of the residuals. Both of them should show no significant non-zero correlations at non-zero lags.

Another method for residual analysis is the tests for white noise. In principle, the Ljung-Box portmanteau test described in the preliminaries section (5) may be applied to test whether residuals are white noise. However, one complication here is that residuals resulting from a fitted model are not raw data. The fact that the data have been used to estimate the parameters in the model leads to the reduction of the degree of freedom in the test. When the Ljung-Box statistics  $Q_m$  defined in (5) is applied for testing the residuals from fitting a stationary ARMA( $p, q$ ) model, the degree of freedom needs to be adjusted from  $\chi_m^2$  to  $\chi_{m-p-q}^2$ .

6) **Model selection:** Although the model diagnostics methods based on residual analysis are effective and widely used in practice, they suffer from an obvious drawback that they are powerless in detecting the problem of overfitting. Overfitting often leads to an unnecessarily complicated model with some redundant parameters, which increases the errors in the estimated parameters. It also makes the interpretation of the fitted model more difficult. Therefore, a good strategy to build a model may be to combine the consideration on both the goodness-of-fit and the simplicity of the model. Akaike's information criterion (AIC) [1] and its numerous variations are proposed to select an optimal model based on the trade-off between those two factors. For a stationary ARMA( $p, q$ ) model with normal innovations, the AIC can be calculated as

$$\text{AIC}(p, q) = T \log \hat{\sigma}^2 + 2(p + q + 1), \quad (11)$$

where  $\hat{\sigma}^2$  is the Gaussian MLE for the variance of innovation defined in (8) [9]. The first term on the RHS of (11) reflects the goodness-of-fit of the model. As  $p$  and  $q$  increase,  $\hat{\sigma}^2$  decreases. The second term on the RHS of (11) is a penalty for the complexity of the model. Obviously, it increases when  $p$  and  $q$  increases. We select the order  $(p, q)$  which minimizes  $\text{AIC}(p, q)$ . AIC, however, does not lead to a consistent order selection [32].

As the AIC tends to overestimate the orders, a popular alternative is the Bayesian information criterion (BIC) [31], which, for a stationary ARMA( $p, q$ ) model with normal innovations, is given by

$$\text{BIC}(p, q) = T \log(\hat{\sigma}^2) + (p + q + 1) \log T. \quad (12)$$

As long as  $\log T > 2$ , BIC penalizes the order more severely than AIC. Hence BIC selects order of ARMA models that are no larger than those by the AIC. Indeed, it can result in oversimplified ARMA models.

It is worth pointing out that the AIC values returned by the R function `arima` are based on a more general formula of AIC, given by

$$\text{AIC} = -2(\text{maximized log-likelihood}) + 2(\text{No. of estimated parameters}). \quad (13)$$

This formula differs from (11) by a common constant. Therefore the difference has no bearings in the model selection.

7) **Forecasting ARMA models:** Let  $Y_1, \dots, Y_T$  be the observations from a time series. Our goal is to forecast the future values  $Y_{T+k}$  for  $k \geq 1$ . Assuming no extra information available, we forecast the future based on the observations  $Y_1, \dots, Y_T$  only. Denote by  $Y_T(k)$  the predictor for  $Y_{T+k}$ . We may seek for the predictor  $Y_T(k) = f_0(Y_1, \dots, Y_T)$  as a function of observed data such that the mean squared predictive error (MSPE)

$$\mathbb{E}[\{Y_{T+k} - f(Y_1, \dots, Y_T)\}^2] \quad (14)$$

is minimized among all predictors  $f$ . Proposition 3.2 of Fan and Yao (2003) [8] indicates that the least squares predictor is

$$Y_T(k) = \mathbb{E}(Y_{T+k} | Y_T, Y_{T-1}, \dots, Y_1), \quad k \geq 1. \quad (15)$$

The best predictor  $Y_T(k)$  is simply the average of all possible realization  $Y_{T+k}$  given the information up to time  $T$ . The rest of the math derivation is rather technical, so we decide to present it in Appendix D.

## B. Holt-Winters' Seasonal Method

1) **Exponential smoothing:** Simple exponential smoothing is a class of forecasting methods proposed in the late 1950s (Holt, 1957 [18]; Winters, 1960 [37]). Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight [20]. While ARIMA is model based, exponential smoothing methods are data-driven.

The simplest of the exponentially smoothing methods is naturally called simple exponential smoothing (SES). This method is suitable for forecasting data with no clear trend or seasonal pattern. Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past — the smallest weights are associated with the oldest observations:

$$\hat{Y}_{T+1|T} = \alpha Y_T + \alpha(1 - \alpha)Y_{T-1} + \alpha(1 - \alpha)^2 Y_{T-2} + \dots \quad (16)$$

where  $0 \leq \alpha \leq 1$  is called the smoothing parameter.  $\alpha$  controls the rate at which the weights decrease.

From (16), we can see that the forecast at time  $T + 1$  is equal to a weighted average between the most recent observation  $Y_T$  and the previous forecast  $\hat{Y}_{T|T-1}$ :

$$\hat{Y}_{T+1|T} = \alpha Y_T + (1 - \alpha)\hat{Y}_{T|T-1}. \quad (17)$$

Similarly, we can write the fitted values as

$$\hat{Y}_{t+1|t} = \alpha Y_t + (1 - \alpha)\hat{Y}_{t|t-1} \quad (18)$$

for  $t = 1, \dots, T$ . This method requires initialization since the forecast for period one requires the forecast at period zero, which we do not (by definition) have. Hence we let the first fitted value at time 1 be denoted by  $\ell_0$  (which we will have to estimate). Then

$$\begin{aligned}\hat{Y}_{2|1} &= \alpha Y_1 + (1 - \alpha)\ell_0 \\ \hat{Y}_{3|2} &= \alpha Y_2 + (1 - \alpha)\hat{Y}_{2|1} \\ &\dots \\ \hat{Y}_{T|T-1} &= \alpha Y_{T-1} + (1 - \alpha)\hat{Y}_{T-1|T-2} \\ \hat{Y}_{T+1|T} &= \alpha Y_T + (1 - \alpha)\hat{Y}_{T|T-1}\end{aligned}$$

Substituting each equation into the following equation, we obtain

$$\hat{Y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1 - \alpha)^j Y_{T-j} + (1 - \alpha)^T \ell_0. \quad (19)$$

The last term becomes tiny for large  $T$ . Hence the weighted average form leads to the same forecast given by (16).

An alternative representation of simple exponential smoothing is the component form, which is given by

$$\text{Forecast equation } \hat{Y}_{t+k|t} = \ell_t \quad (20)$$

$$\text{Smoothing equation } \ell_t = \alpha Y_t + (1 - \alpha)\ell_{t-1} \quad (21)$$

where  $\ell_t$  is called the level (or the smoothed value) of the series at time  $t$ . Setting  $k = 1$  gives the fitted values, while setting  $t = T$  gives the true forecasts beyond the training data.

The forecast equation shows that the forecast value at time  $t + 1$  is the estimated level at time  $t$ . The smoothing equation for the level (usually referred to as the level equation) gives the estimated level of the series at each period  $t$  [23]. The component form of simple exponential smoothing is not particularly useful, but it will be the easiest form to use when we start adding other components.

The unknown parameters and the initial values for any exponential smoothing methods can be estimated by least squares, i.e. we want and initial values that minimize

$$\text{SSE} = \sum_{t=1}^T (Y_t - \hat{Y}_{t|t-1})^2 = \sum_{t=1}^T e_t^2. \quad (22)$$

For the case of simple exponential smoothing, There are two parameters to estimate:  $\alpha$  and  $\ell_0$ . This is a non-linear minimisation problem, and we need to use an optimisation tool to solve it [23].

Simple exponential smoothing has a "flat" forecast function:

$$\hat{Y}_{T+k|T} = \hat{Y}_{T+1|T} = \ell_T, \quad k = 2, 3, \dots \quad (23)$$

In other words, all forecasts take the same value, which is equal to the last level component. This method may sound trivial, but it serves as a building block for more advanced smoothing methods.

2) **Holt's linear trend method:** If the data have no trend or seasonal patterns, then simple exponential smoothing is appropriate. If the data exhibit a linear trend, then Holt's linear method (or the damped method) is appropriate.

Holt's linear method [18] extends simple exponential smoothing with a trend component. This method involves a forecast equation and two smoothing equations; one for the level and one for the trend:

$$\text{Forecast equation } \hat{Y}_{t+k|t} = \ell_t + kF_t \quad (24)$$

$$\text{Smoothing equation } \ell_t = \alpha Y_t + (1 - \alpha)(\ell_{t-1} + F_{t-1}) \quad (25)$$

$$\text{Trend equation } F_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)F_{t-1}, \quad (26)$$

where  $\ell_t$  denotes an estimate of the level of the series at time  $t$ ,  $F_t$  denotes an estimate of the trend of the series at time  $t$ ,  $\alpha$  is the smoothing parameter for the level,  $0 \leq \alpha \leq 1$ , and  $\beta$  is the smoothing parameter for the trend,  $0 \leq \beta \leq 1$ . In the special case where  $\alpha = \beta$ , Holt's method is equivalent to "Brown's double exponential smoothing" [4].

The level equation here shows that  $\ell_t$  is a weighted average of observation  $Y_t$  and the one-step ahead training forecast for time  $t$ , which is given by  $\ell_{t-1} + F_{t-1}$ . The trend equation shows that  $F_t$  is a weighted average of the estimated trend at time  $t$  based on  $\ell_t - \ell_{t-1}$  and the previous estimate of the trend  $F_{t-1}$ .

The forecast function now is no longer flat but trending. The  $k$ -step ahead forecast is equal to the last estimated level plus  $k$  times the last estimated trend value. Hence the forecasts are a linear function of  $k$ .

The forecasts generated by Holt's linear method display a constant trend (increasing or decreasing) indefinitely into the future. Empirical evidence indicates that these methods tend to over-forecast, especially for longer forecast horizons [23]. Gardner and McKenzie [11] proposed a modification of Holt's linear method to allow the "damping" of trends. This modification is explained in Appendix E.

3) **Holt-Winters' seasonal method:** Holt [18] and Winters [37] extended Holt's linear method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the level  $\ell_t$ , one for the trend  $F_t$ , and one for the seasonal component  $G_t$ , with corresponding smoothing parameters  $\alpha$ ,  $\beta$ , and  $\gamma$ . We use  $m$  to denote the period of the seasonality, i.e., the number of seasons in a year. For example, for quarterly data  $m = 4$ , for monthly data  $m = 12$ , for daily data  $m = 7$ .

There are two different Holt-Winters' methods, depending on whether seasonality is modeled in an additive or multiplicative way. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

The component form for the additive method is:

$$\text{Forecast equation } \hat{Y}_{t+k|t} = \ell_t + kF_t + G_{t+m+k_m^+} \quad (27)$$

$$\text{Smoothing equation } \ell_t = \alpha(Y_t - G_{t-m}) + (1 - \alpha)(\ell_{t-1} + F_{t-1}) \quad (28)$$

$$\text{Trend equation } F_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)F_{t-1} \quad (29)$$

$$\text{Seasonal equation } G_t = \gamma(Y_t - \ell_{t-1} - F_{t-1}) + (1 - \gamma)G_{t-m}, \quad (30)$$

where  $k_m^+ = [(k - 1) \bmod m] + 1$ . The parameters  $(\alpha, \beta, \gamma)$  are usually restricted to lie between 0 and 1.

The component form for the multiplicative method is:

$$\text{Forecast equation } \hat{Y}_{t+k|t} = (\ell_t + kF_t) \cdot G_{t+m+k_m^+} \quad (31)$$

$$\text{Smoothing equation } \ell_t = \alpha \frac{Y_t}{G_{t-m}} + (1 - \alpha)(\ell_{t-1} + F_{t-1}) \quad (32)$$

$$\text{Trend equation } F_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)F_{t-1} \quad (33)$$

$$\text{Seasonal equation } G_t = \gamma \frac{Y_t}{(\ell_{t-1} + F_{t-1})} + (1 - \gamma)G_{t-m}. \quad (34)$$

### C. Vector Autoregressive (VAR) Models

1) **Vector autoregressive (VAR) models:** If needed, preliminaries and notation of time series vectors are included in Appendix A, concepts of nonstationary time series and strategies for multivariate time series analysis are gently introduced in Appendix F and G.

In the ARDL models (92), we assumed that one of the variables was the dependent variable and the other was the independent variable, and we treated the relationship between  $Y_t$  and  $X_t$  like a regression model. However, a priori, unless we have good reasons not to, we could have assumed that  $Y_t$  is the independent variable and  $X_t$  is the dependent variable [17]. A *vector autoregressive (VAR)* model is useful when one is interested in predicting multiple time series variables using a single model. At its core, the VAR model is an extension of the univariate autoregressive model.

A  $d$ -vector autoregressive model with order  $p$  is of the form

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (35)$$

where  $\boldsymbol{\varepsilon}_t \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$  is a  $d \times 1$  vector,  $\mathbf{Y}_t$  is a  $d \times 1$  vector,  $\mathbf{c}$  is a  $d \times 1$  constant vector, and  $\mathbf{A}_1, \dots, \mathbf{A}_p$  are  $d \times d$  autoregressive coefficient matrices. Note that vectors are always in column in our notation. We write  $\mathbf{Y}_t \sim \text{VAR}(p)$  to emphasize the vector part of the AR model. In this model, each component of  $\mathbf{Y}_t$  is a linear combination of its lagged values and the lagged values of other components. As an example, suppose we have two time series  $\{Y_t, X_t\}$  which are both I(1) nonstationary and are not cointegrated, then as discuss in previous section, we work with the first difference. In this case, a VAR(1) model would be

$$\begin{aligned} \nabla Y_t &= A_{11} \nabla Y_{t-1} + A_{12} \nabla X_{t-1} + \varepsilon_t^{\nabla Y} \\ \nabla X_t &= A_{21} \nabla Y_{t-1} + A_{22} \nabla X_{t-1} + \varepsilon_t^{\nabla X}. \end{aligned} \quad (36)$$

Note that all variables are now I(0), so the system can be estimated by least squares.

2) **Fitting VAR models:** Estimation for vector autoregressive models is similar to that for univariate models: it can be done by one of the three methods: least squares method, Yule-Walker estimation (i.e. method of moments estimation), and quasi maximum likelihood estimation based on, for example, Gaussian innovation distribution. The estimator obtained by all these three methods are asymptotically normal and also asymptotically equivalent under some mild conditions [9]. Perhaps the least squares estimation (LSE) is the easiest way in this context. Therefore we will only present LSE in this paper. To avoid cumbersome notation, we illustrate below the LSE method for a VAR(2) model.

Let  $\mathbf{Y}_1, \dots, \mathbf{Y}_T$  be observations from

$$\mathbf{Y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \mathbf{A}_2 \mathbf{Y}_{t-2} + \boldsymbol{\varepsilon}_t, \quad (37)$$

where  $\boldsymbol{\varepsilon}_t \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$ . The goal is to estimate the parameters  $\mathbf{c}, \mathbf{A}_1, \mathbf{A}_2, \Sigma_\varepsilon$ . Note that the parameters in  $\mathbf{c}, \mathbf{A}_1, \mathbf{A}_2$  can be decoupled row by row. For example, the equation for the first component under model (37) can be written as

$$Y_{t1} = c_1 + \mathbf{Y}_{t-1}^\top \mathbf{a}_1^{(1)} + \mathbf{Y}_{t-2}^\top \mathbf{a}_1^{(2)} + \varepsilon_{t1}, \quad (38)$$

where  $c_1$  is the first component of  $\mathbf{c}$ , and  $\mathbf{a}_i^{(j)}$  is the  $i$ th row vector of the matrix  $\mathbf{A}_j$ . Thus, the parameters associated with the first time series can be estimated by the least squares method [9], which minimizes the sum of squared fitting errors with respect to  $c_1, \mathbf{a}_1^{(1)}, \mathbf{a}_1^{(2)}$ :

$$\sum_{t=3}^T (Y_{t1} - c_1 - \mathbf{A}_1 \mathbf{Y}_{t-1} - \mathbf{A}_2 \mathbf{Y}_{t-2})^2. \quad (39)$$

Applying the above estimation method to each component of (37), we obtain the estimators  $\hat{\mathbf{c}}, \hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2$ . Then the estimator for  $\Sigma_\varepsilon$  can be defined as [9]

$$\hat{\Sigma}_\varepsilon = \frac{1}{T-2} \sum_{t=3}^T \hat{\varepsilon}_t \hat{\varepsilon}_t^\top, \quad (40)$$

where

$$\hat{\varepsilon}_t = \mathbf{Y}_t - \hat{\mathbf{A}}_1 \mathbf{Y}_{t-1} - \hat{\mathbf{A}}_2 \mathbf{Y}_{t-2}.$$

3) **Model diagnostics:** The most frequently used method for model diagnostic is to examine if the residuals

$$\hat{\varepsilon} = \mathbf{Y}_t - \hat{\mathbf{c}} - \sum_{j=1}^p \hat{\mathbf{A}}_j \mathbf{Y}_{t-j}, \quad t = p+1, \dots, T \quad (41)$$

behave like vector white noise. Let  $\hat{\Gamma}(k)$  denote the sample cross variance matrix of the residuals at lag  $k$ . Then the portmanteau test statistics is defined as [9]

$$Q_m = T^2 \sum_{j=1}^m \frac{1}{T-j} \text{tr}\{\hat{\Gamma}(k)^\top \hat{\Gamma}(0)^{-1} \hat{\Gamma}(k) \hat{\Gamma}(0)^{-1}\}, \quad (42)$$

where  $m > p$  is an integer. If the true model is indeed  $\text{VAR}(p)$  with  $\varepsilon_t \sim \text{IID}(0, \Sigma_\varepsilon)$ ,  $Q_m$  is asymptotically  $\chi_{d^2(m-p)}^2$  distributed [9]. This is an extension of the Ljung-Box test in preliminareis (5). We compute the p-value by computing the right tail probability of  $\chi_{d^2(m-p)}^2$  distribution at observed test statistic  $Q_m$ .

4) **Model selection:** In fitting an VAR model, an important step is to determine the order  $p$ . This can be done either via hypothesis testing or in terms of some information criteria.

To determine if a model with order  $p > 1$  is sufficiently large, we may test the hypotheses:

$$\begin{aligned} H_0 : \mathbf{Y}_t &= \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \dots + \mathbf{A}_p \mathbf{Y}_{t-p} + \varepsilon_t, \quad \text{against} \\ H_1 : \mathbf{Y}_t &= \mathbf{c} + \mathbf{A}_1 \mathbf{Y}_{t-1} + \dots + \mathbf{A}_{p+1} \mathbf{Y}_{t-p-1} + \varepsilon_t, \end{aligned}$$

where  $\varepsilon_t \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$ . Under the  $\text{VAR}(p)$  model above, define estimator for  $\Sigma_\varepsilon$  as

$$\hat{\Sigma}_\varepsilon(p) = \frac{1}{T-2p-1} \sum_{t=p+1}^T \left( \mathbf{Y}_t - \hat{\mathbf{c}} - \sum_{j=1}^p \hat{\mathbf{A}}_j \mathbf{Y}_{t-j} \right) \left( \mathbf{Y}_t - \hat{\mathbf{c}} - \sum_{j=1}^p \hat{\mathbf{A}}_j \mathbf{Y}_{t-j} \right)^\top, \quad (43)$$

where  $\hat{\mathbf{c}}$  and  $\hat{\mathbf{A}}_j$  are the least squares estimates for  $\mathbf{c}$  and  $\mathbf{A}_j$ , respectively. The test statistic for the hypotheses above is defined as [9]

$$S = (T-d-p-1/2) \log(|\hat{\Sigma}_\varepsilon(p)|/|\hat{\Sigma}_\varepsilon(p+1)|). \quad (44)$$

We reject  $H_0$  in favor of  $H_1$  for large values of  $S$ . Under  $H_0$ ,  $S$  is asymptotically  $\chi^2$ -distributed with  $d^2$  degrees of freedom [34].

Alternatively we can also use an information criterion to select the order  $p$ . For a  $\text{VAR}(p)$  model, the AIC and BIC are defined respectively as [9]

$$\text{AIC}(p) = \log(|\hat{\Sigma}_\varepsilon(p)|) + 2d^2 p/T, \quad (45)$$

$$\text{BIC}(p) = \log(|\hat{\Sigma}_\varepsilon(p)|) + \log(T)d^2 p/T. \quad (46)$$

We choose  $p$  such that one of these criteria is minimized.

5) **Granger causality:** *Granger causality* [13] refers to the ability of lags of one variable to contribute to the forecast of another variable. Let  $Z_t$  and  $X_t$  be two univariate time series., time series  $Z_t$  is said to Granger cause time  $X_t$  if

$$\mathbb{E}(X_t | X_{t-1}, Z_{t-1}, X_{t-2}, Z_{t-2}, \dots) \neq \mathbb{E}(X_t | X_{t-1}, X_{t-2}, \dots). \quad (47)$$

The causality (47) can easily be verified for the VAR models. Let  $\mathbf{Y}_t = (Z_t, X_t)^\top$ , and assume  $\mathbf{Y}_t \sim \text{VAR}(p)$ , i.e.

$$\mathbf{Y}_t = \mathbf{c} + \sum_{j=1}^p \mathbf{A}_j \mathbf{Y}_{t-j} + \varepsilon_t, \quad (48)$$

where  $\varepsilon_t \sim \text{WN}(\mathbf{0}, \Sigma_\varepsilon)$ . Under the additional assumption that  $\mathbf{Y}_t$  is a Gaussian process, the conditional expectation in (47) are linear in the variables that are conditioned upon. The first conditional expectation is determined by the second component of the vector  $\mathbf{c} + \sum_{j=1}^p \mathbf{A}_j \mathbf{Y}_{t-j}$ , which depends on some lags of  $Z$ 's unless  $a_{21}^{(1)} = \dots = a_{21}^{(p)} = 0$ , where  $a_{21}^{(k)}$  denotes the (2,1)-element of the coefficient matrix  $\mathbf{A}_k$  in (48). Therefore, (47) is equivalent to the condition that at least one of  $a_{21}^{(k)} \neq 0$  for  $1 \leq k \leq p$ . This so-called *Granger causality test* is to test the hypothesis

$$H_0 : a_{21}^{(1)} = \dots = a_{21}^{(p)} = 0$$

for model (48). When  $H_0$  is rejected,  $Z_t$  is regarded as Granger causing  $X_t$ .

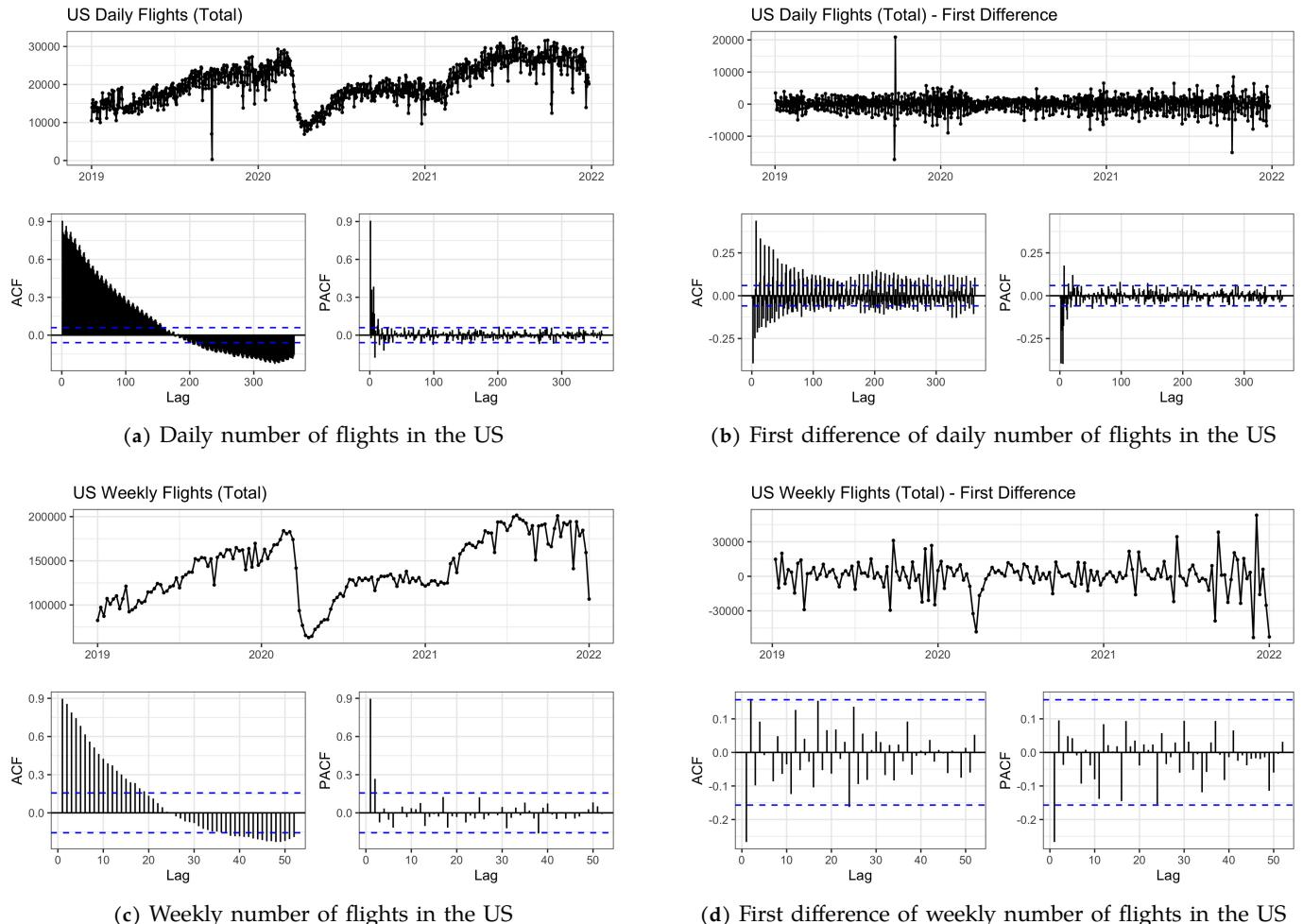
6) **Impulse response functions (IRF):** In addition to the Granger causality, another way to investigate the effect of a change in one component series on the other components is via *impulse response functions*. For example, in macroeconomics, this technique can be used to analyze problems such as the effect of an oil price shock on inflation and GDP growth, and the effect of a change in monetary policy on the economy. Although the mathematics behind this method can be technical, the interpretation is straightforward. We therefore will not cover the math derivation in this paper; see §11.4 of James D. Hamilton (1994) [16].

#### IV. FORECASTING NUMBER OF FLIGHTS - RESULTS

##### A. Forecasts from ARIMA Models

1) **Stationarity:** Figure 8a shows the time series of daily number of flights in the US (sum of domestic, international arrival, and international departure), along with its ACF and PACF plots. From the time series plot it self, it is hard to tell if the variable is stationary or not. When we run the augmented Dickey-Fuller test (68) on this time series, we get p-value = 0.2997. Thus US daily flights is not a stationary time series. Figure 8b shows the first difference of US daily flights, along with its ACF and PACF plots. The series shows some mean-reverting behavior, and the ACF and PACF plots also indicate that it is stationary. When we run the augmented Dickey-Fuller test (67), we get p-value = 0.01. Hence the first difference is stationary.

Figure 8c and 8d show the weekly counterparts for Figure 8a and 8b. The p-values from the augmented Dickey-Fuller tests are 0.3094 and 0.01, respectfully, indicating that US weekly flights is a nonstationary I(1) variable, i.e., its first difference is stationary.



**Figure 8:** Time series, ACF and PACF plots of daily and weekly number of flights in the US and their first differences

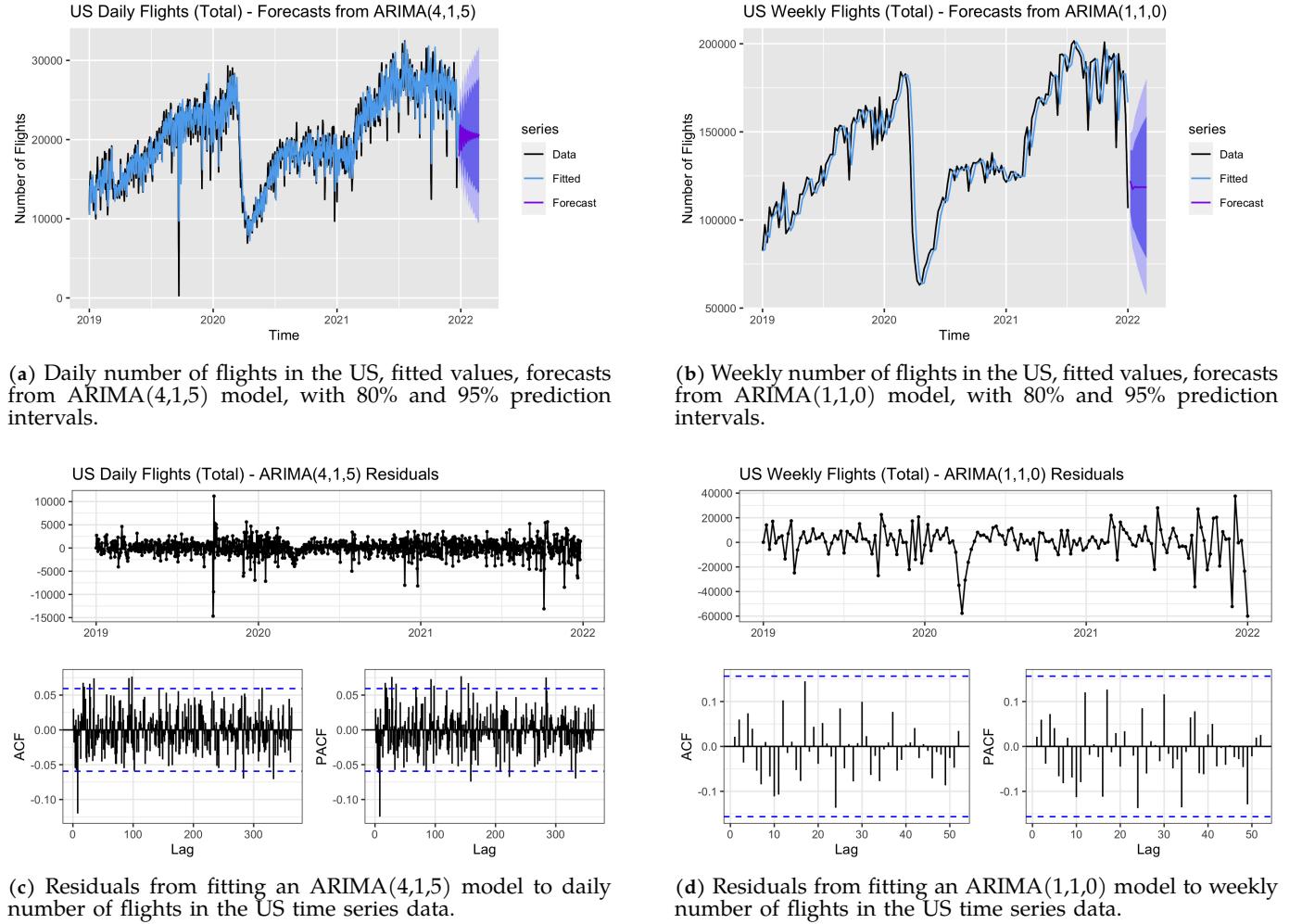
Therefore, it is clear that we can apply stationary ARMA( $p, q$ ) models to the first difference of either daily or weekly data. In other words, if we want to fit an ARIMA( $p, d, q$ ) model to the original series, the parameter  $d$  should be equal to 1 for both daily and weekly data.

2) **Forecasting results:** We have shown that the parameter  $d$  for both daily and weekly models should be 1. Now we move on to select the best order ( $p, q$ ) for each model. For daily data,  $(p = 4, q = 5)$  minimizes both AIC (11) and BIC (12). For weekly data, the best combination is  $(p = 1, q = 0)$ . This result is consistent with the PACF plot of the first difference of weekly number of flights in the US, as we can observe one spike at lag 1. Therefore, we will fit ARMA(4, 5) to the first difference of the daily number of flights and ARMA(1, 0) to the first difference of the weekly series. This is equivalent to fitting an ARIMA(4, 1, 5) model to the daily data itself and an ARIMA(1, 1, 0) model to the weekly data.

Figure 9a shows the forecasts of daily number of US flights from ARIMA(4,1,5) model with an 80% and 90% prediction interval. Figure 9b shows a similar graph except this is for forecasts of weekly number of flights in the US from ARIMA(1,1,0) model. It is clear on both graphs that the forecasts show a very large range of possible values, although even the most optimistic values do not reach the peak flight numbers from a couple months prior.

3) **Residual analysis:** Figure 9c shows the residuals and ACF and PACF plots from fitting the ARIMA(4,1,5) model on daily number of US flights. When applying the Ljung-Box test, we get a p-value of 0.3138, meaning the residuals are white noise

for this model. Figure 9d shows similar information as before except this is applied on the ARIMA(1,1,0) model on weekly number of US flights. When applying the Ljung-Box test here, we get a p-value of 0.7851, meaning the residuals are white noise for this model. Thus for both models, the autocorrelations of the residuals are very small and the models don't show a significant lack of fit.



**Figure 9:** Forecasts from ARIMA models and residual plots.

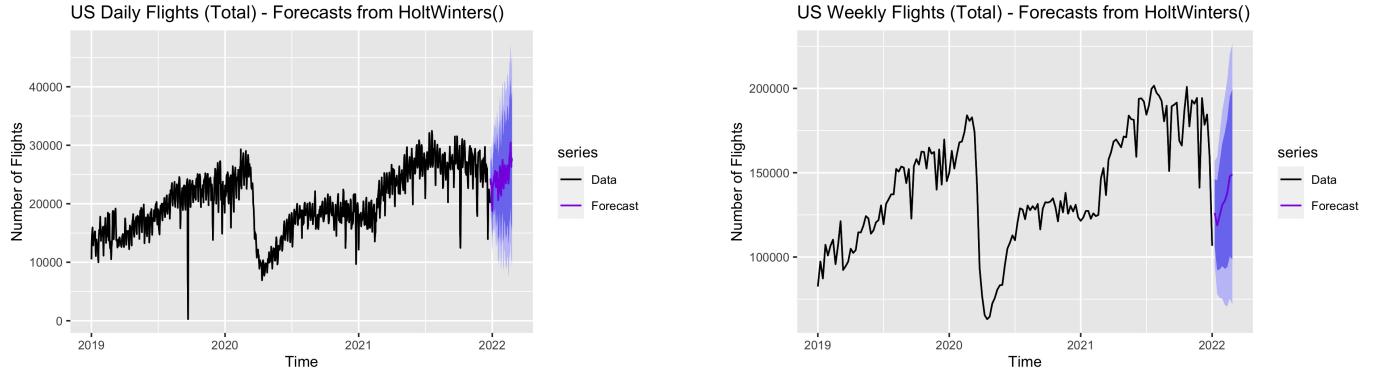
### B. Forecasts from Holt-Winters' Method

1) **Implementation:** There are two different implementations of the Holt Winters' method in R, namely, `HoltWinters()` in base R and `hw()` in the `forecast` package. `hw()` is a wrapper function of `ets()` (exponential smoothing state space model).

According to Rob J. Hyndman, the author of the `forecast` package [21], there are several differences involved. First of all, `HoltWinters()` and `ets()` are optimizing different criterion. `HoltWinters()` is using heuristic values for the initial states and then estimating the smoothing parameters by optimizing the MSE. `ets()` is estimating both the initial states and smoothing parameters by optimizing the likelihood function (which is only equivalent to optimizing the MSE for the linear additive models). Second, the two functions use different optimization routines and different starting values. That wouldn't matter if the surfaces being optimized were smooth, but they are not. Because the MSE and likelihood surfaces are both fairly bumpy, it is easy to find a local optimum. The only way to avoid this problem is to use a much slower computational method such as PSO. Thirdly, `ets()` searches over a restricted parameter space to ensure the resulting model is forecastable. `HoltWinters()` ignores this issue (it was written before the problem was even discovered). In what follows, we use both implementations to make forecasts.

2) **Forecasting results from `stats::HoltWinters()`:** Figures 10a and 10b result from using `HoltWinters()` in base R, which allows us to apply the method on daily and weekly number of flights in the US. Notice that in both graphs, there is a clear upward trend of the forecast values. This indicates that according to this method, flight numbers will start to reach previous values relatively quickly.

We highly suspect this implementation of overfitting our data because for daily data  $m = 365$ , we need to estimate  $(m-1) = 364$  initial states. Similarly, we need to estimate 51 initial states for weekly data. For large  $m$  like this, the estimation becomes almost impossible.



(a) HoltWinters(): daily number of flights in the US

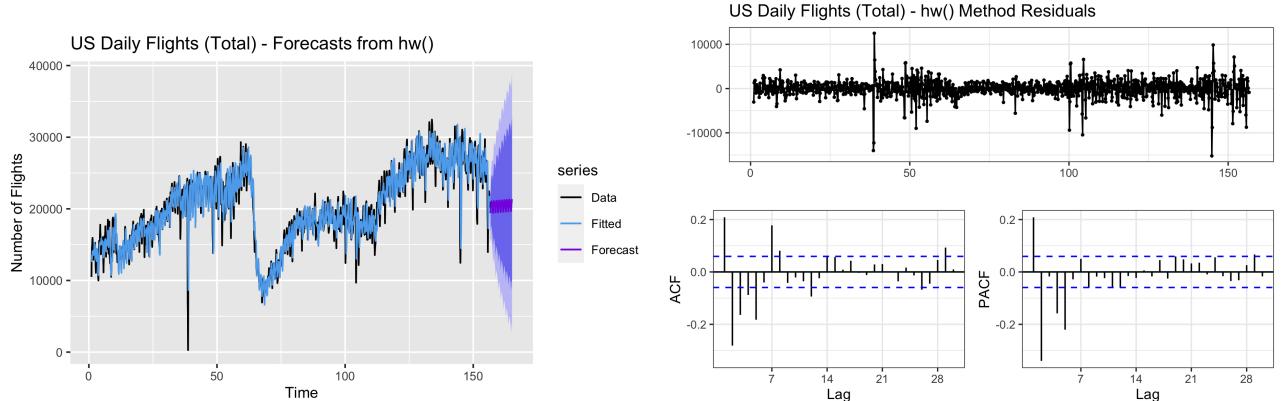
(b) HoltWinters(): weekly number of flights in the US

Figure 10: Forecasts from HoltWinters() and hw() functions, with 80% and 95% prediction intervals.

3) **Forecasting results from `forecast`: `:hw()`:** Rob Hyndman's suggestion [21] is to use the `hw()` function in the `forecast` package with specified  $m = 7$  argument for daily data.

Figure 11a results from using the `hw()` method from the `forecast` package. The downside of using this package is that we can only apply it to daily number of flights in the US. The forecast in this graph does not look like it is increasing or decreasing, in contrast to the graphs using the `HoltWinters()` method.

Figure 11b shows the residuals from fitting `hw()` to the daily number of flights in the US. We observe some spikes in both ACF and PACF plots of the residual. Hence we suspect there exists some serial correlation in the residuals. Box-Ljung test (5) shows a p-value smaller than 5.375e-12 when we test the residuals against the null hypothesis of white noise, so the autocorrelations of the residuals are not very small. Therefore, the Holt Winters' method might not able to capture the dynamic nature of the time series.



(a) hw(): daily number of flights in the US

(b) Residuals

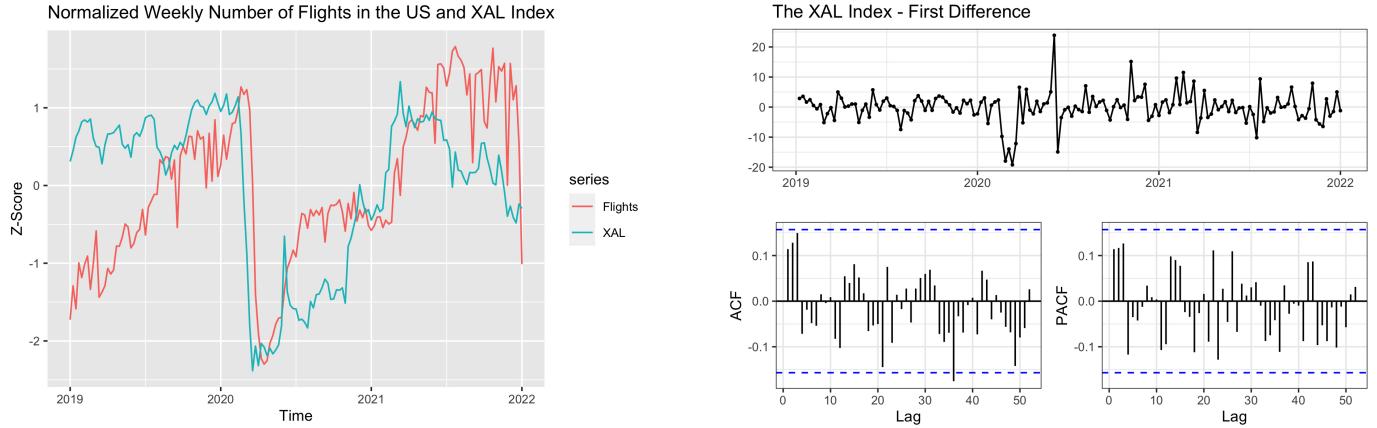
Figure 11: Forecasts from hw() functions, with 80% and 95% prediction intervals.

### C. Forecasts from VAR Models

1) **Stationarity:** The weekly COVID confirmed cases in the US from Jan. 2020 to Dec. 2021 is more than I(4) nonstationary and our ADF test (68) fails to reject the null for the fourth difference of the series. Therefore based on our discussion in Section III-C we should not include the COVID confirmed cases series in our VAR model.

Instead, we decide to fit a VAR model on XAL, the NYSE ARCA Airline Index, and the weekly number of flights. We think this index is useful because we believe the stock market and the stock price reflect all the information about this industry, e.g. investors' expectation of COVID, number of flights, etc.

As mentioned before, the weekly number of flights in the US is a I(1) nonsationary time series. The XAL index is I(1) nonstationary as well. When we apply our ADF test (68) to the XAL index, we get p-value = 0.5777, but the p-value for the first difference of the XAL index is less than 0.01, meaning the first different of XAL is stationary.



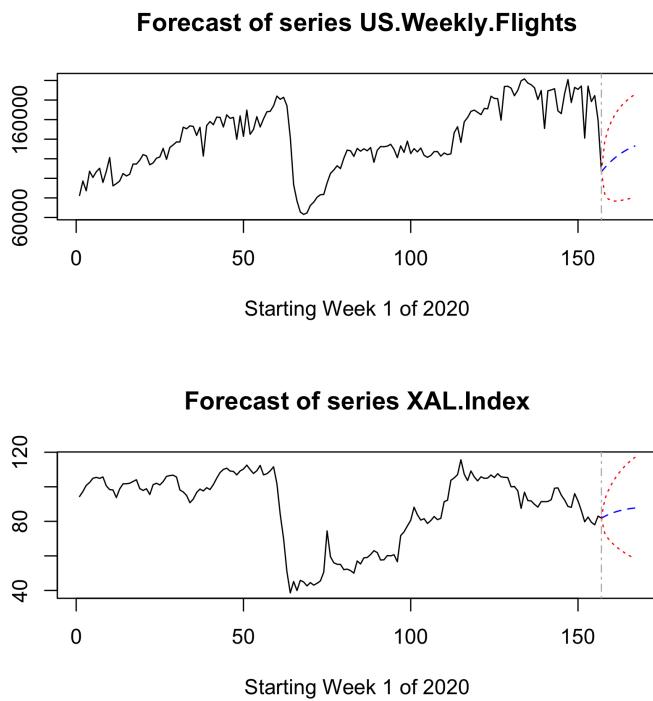
(a) Normalized XAL and US weekly number of flights.

(b) First difference of XAL.

**Figure 12:** Both the XAL index and the US weekly number of flights are I(1) nonstationary.

2) *Cointegration:* We use two methods to test if XAL and US weekly number of flights are cointegrated. We first run ADF test on the difference of the normalized values (subtract mean and divide by standard deviation) of the two series, which gives us a p-value of 0.2499, meaning the difference is not I(0) stationary, i.e. the series are not cointegrated. We also run the Phillips-Ouliaris cointegration test (`tseries::po.test()`) and get a p-value of 0.1139, which means we cannot reject the null hypothesis that these two series are not cointegrated. Therefore, we believe XAL and US weekly number of flights are not cointegrated. According to Figure 15, we should estimate a VAR model using the first difference of each series.

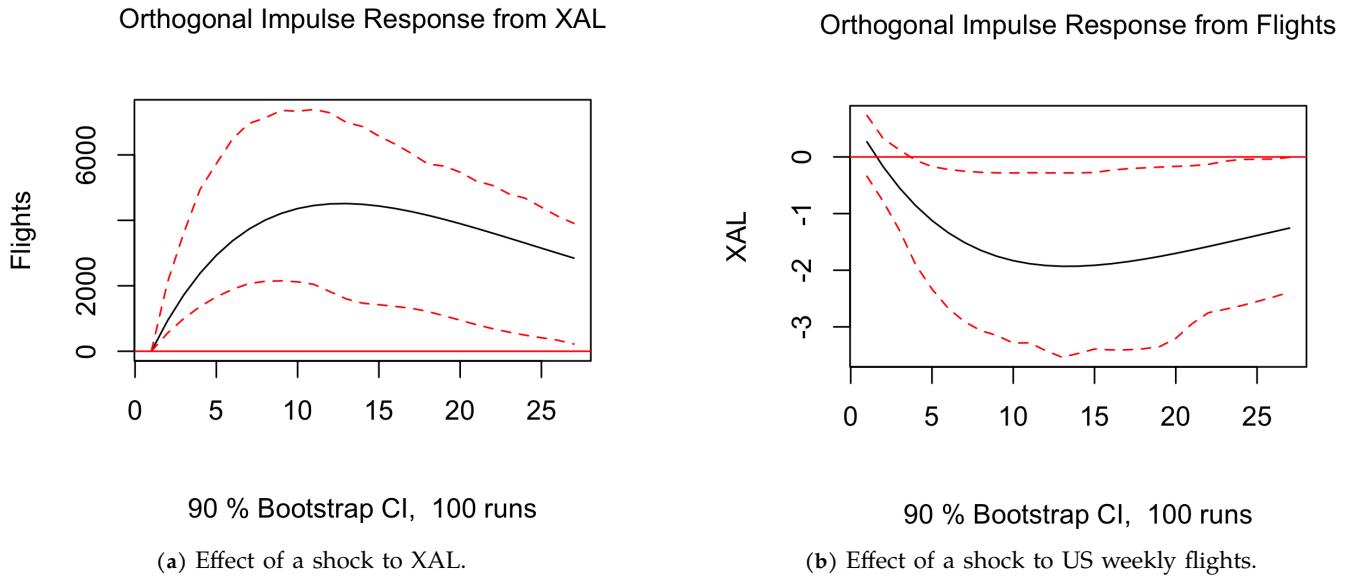
3) *Forecasting results:* Now we need to select the best order  $p$  which minimizes the AIC or BIC. Using the R function `VARselect()` in the `vars` package, AIC is minimized when  $p = 6$ , and BIC is minimized when  $p = 1$ . Seeking a parsimonious model, we choose  $p = 1$ .

**Figure 13:** Forecast of US Weekly Flights and the XAL Index. The graph starts at Week 1 of 2020, while the forecast starts at Week 1 of 2022

Both graphs show a similar story where there is a big drop during the initial outbreak of the virus. Surprisingly, both plots reach pre-Covid numbers rather quickly. The graph of flights consistently stayed at a higher point than the XAL Index, but the end of 2021 saw another big drop for the US Weekly Flights graph. The XAL Index plot started to decrease earlier than the flights graph, but it did not have as big of a drop near the end of 2021. Looking at the forecasts, US Weekly Flights predicts an increase while the XAL Index slightly increases.

4) *Causality*: We run Granger causality test (R function `grangertest()`) on both directions. In the first Granger causality test, we test the effect of XAL on forecasting US weekly number of flights. The resulting p-value of 0.00176 shows that lags of XAL series has a strong ability to contribute to the forecast of weekly number of flights in the US. In the second Granger causality test, we test the other way around. We get a small p-value of 0.01633, which says that the lags of US weekly number of flights have some ability to forecast the XAL index. This is consistent with the visualization in Figure 12a, especially during the initial outbreak of COVID-19 in February to March 2020, as the blue curve “triggers” the movements of the red curve.

5) *IRF*: Figure 14 visualizes the impulse response functions of each time series. In each graph, if zero is within the 95% confidence bands (the red dotted lines), then we will consider that the response is statistically zero. The left plot tells us that a positive shock to XAL will cause a significant increase in weekly number of flights in the US, and the effect of the shock persists for about 26 weeks. Similarly, the right plot tells us that a positive shock to US weekly flights will cause a significant decrease in the XAL index, and the effect of the shock persists for about half an year as well.



**Figure 14:** Impulse response function plots.

## V. CONCLUSIONS

### A. Conclusions

No company or business went unscathed by the onslaught of the pandemic; however, few areas were hurt as bad as the airline industry. From Section II of our project, it is evident that COVID-19 had a significant impact on flight numbers around the world and affected airline revenue immensely. When looking at the US specifically, it is clear that there is a negative correlation between weekly number of flights and confirmed cases. No matter how we analyzed it, from splitting passengers into domestic and international to visualizing number of US flights in a network, we saw that there was a significant decrease in passengers and flights during the outbreak of the virus. From Section IV of our project, we see how the future of flights could look. From the ARIMA models, we see that the forecasts for daily and weekly flights seem to be constant over the long run. This is consistent with the nature of ARIMA models, where making long-term forecasting is difficult and the best predictor converges to the unconditional mean of the time series. The Holt-Winters method saw interesting results, as the base R `HoltWinters()` function predicted an increase in flight numbers while the `hw()` function from the forecast package predicted constant values over the long run. This is most likely due to the different approaches for choosing the initial states and parameter values. However, the residuals plot of the `hw()` method did not pass the Box-Ljung test, meaning we do not want to put too much trust into that forecast. The VAR models looked at US weekly flights and the XAL Index. These forecasts both showed an increase in predicted values, although the weekly flights plot had a sharper increase than the XAL Index graph. Given that the Granger causality test resulted in very low p-values, we can say that these are very strong predictions.

### B. Ethical Implications

The ethical use of data can be a tricky subject. The internet allows us to obtain endless amounts of information, but this information is not always our own to use. There can be issues of privacy, exploitation, confidentiality, and more if one is not careful. However, all data in this project is available publicly and keeps any sensitive information private. According to [6], there are five principles of data ethics. The first principle is ownership, and since all the data we used is collected publicly as a group and not individually, we are not using personal data without consent. Since our sources are widely used public data sets, any information on these sites are already checked for security issues. The second principle is transparency, which involves saying how we plan to use the data. We plainly state what kind of forecasting methods we use, and we show all plots using these methods so there is no question of how we use this data. We do not withhold any results we obtain, so anything good or bad is shown clearly. Privacy is another concern, but none of the sources we used included any personally identifiable information. All of our sources are de-identified, meaning all that is left is anonymous data. As far as intention goes, the results of our project are not harming anyone and we are not using this for our own personal gain. Also, any data we use is not sensitive nor

unethically collected. We are using these results purely to understand the impact of COVID-19 on the flight industry. The final principle is the outcome of our data analysis. Our project does not look at the individual at all. We only compare groups at the country level, and that involves total flight numbers and COVID cases per country. The forecasting results we obtained are only predictions and should not be used as a concrete guide of the future of the airline industry. Any conclusions we make are a direct result of our analysis and do not include our own opinion.

## APPENDIX

### A. Time Series Preliminaries and Notation - Vector Version

Let  $\mathbf{Y}_t = (Y_{t1}, \dots, Y_{td})^\top$  be  $d$  time series, where  $Y_{tj}$  is the  $j$ th component series.  $\mathbf{Y}_t$  is *weakly stationary* or simply *stationary*, if all its first and second moments are time-invariant, i.e.

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{Y}_t), \quad \boldsymbol{\Gamma}(k) = \mathbb{E}\{(\mathbf{Y}_{t+k} - \boldsymbol{\mu})(\mathbf{Y}_t - \boldsymbol{\mu})^\top\} \quad (49)$$

are independent of  $t$ . The matrix valued function  $\boldsymbol{\Gamma}(\cdot)$  is called the autocovariance matrix function, or *cross covariance* function. Let  $\mu_j$  be the  $j$ th element of  $\boldsymbol{\mu}$ , and  $\gamma_{ij}(k)$  be the  $(i, j)$ th element of  $\boldsymbol{\Gamma}(k)$ . Then

$$\gamma_{ij}(k) = \text{cov}(Y_{t+k,i}, Y_{tj}) = \text{cov}\{(Y_{t+k,i} - \mu_i)(Y_{tj} - \mu_j)\} \quad (50)$$

is called the cross covariance between the  $i$ th and  $j$ th component series at time lag  $k$  for  $i \neq j$ . The autocorrelation matrix, which is also called the *cross correlation matrix*, of  $\mathbf{Y}_t$  is defined as

$$\mathbf{R}(k) = (\rho_{ij}(k)) = \mathbf{D}^{-1/2} \boldsymbol{\Gamma}(k) \mathbf{D}^{-1/2}, \quad (51)$$

where  $\mathbf{D}^{-1/2}$  is the diagonal matrix with  $\gamma_{jj}(0)^{-1/2}$ , the reciprocal of the standard deviation of the  $j$ th series, as its  $j$ th main diagonal entry.

With available observations  $\mathbf{Y}_1, \dots, \mathbf{Y}_T$  from a weakly stationary process, a natural estimator of the cross covariance matrix is the *sample cross covariance matrix*

$$\widehat{\boldsymbol{\Gamma}}(k) = (\widehat{\gamma}_{ij}(k)) = \frac{1}{T} \sum_{t=1}^{T-k} (\mathbf{Y}_{t+k} - \widehat{\boldsymbol{\mu}})(\mathbf{Y}_t - \widehat{\boldsymbol{\mu}})^\top, \quad \widehat{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \mathbf{Y}_t. \quad (52)$$

The cross correlation matrix can be estimated by the *sample cross correlation matrix*

$$\widehat{\mathbf{R}}(k) = (\widehat{\rho}_{ij}(k)) = \widehat{\mathbf{D}}^{-1/2} \widehat{\boldsymbol{\Gamma}}(k) \widehat{\mathbf{D}}^{-1/2}, \quad (53)$$

where  $\widehat{\mathbf{D}} = \text{diag}(\widehat{\gamma}_{11}(0), \dots, \widehat{\gamma}_{dd}(0))$ .

The simplest type of weakly stationary vector processes is *vector white noise* denoted by  $\text{WN}(\mathbf{a}, \boldsymbol{\Sigma}_\varepsilon)$ . We say  $\varepsilon \sim \text{WN}(\mathbf{a}, \boldsymbol{\Sigma}_\varepsilon)$  if  $\mathbb{E}(\varepsilon_t) = \mathbf{a}$ ,  $\text{var}(\varepsilon_t) = \boldsymbol{\Sigma}_\varepsilon$ , and  $\text{cov}(\varepsilon_t, \varepsilon_s) = \mathbf{0}$  for any  $t \neq s$ . Hence there exists no serial correlation across all the components of  $\varepsilon_t$ . However, different components of  $\varepsilon_t$  may be correlated with each other contemporaneously as  $\boldsymbol{\Sigma}_\varepsilon$  is not necessarily a diagonal matrix. Most frequently used white noise is the one with mean  $\mathbf{0}$ , i.e.  $\text{WN}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$ .

### B. Properties of MA and AR Models

Moving average (MA) processes are useful in describing phenomena in which events produce an immediate effect that only lasts for short periods of time. Let  $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ . For a fixed integer  $q \geq 1$ , we write  $Y_t \sim \text{MA}(q)$  if  $Y_t$  is defined as a moving average of  $q$  successive  $\varepsilon$  as follows:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (54)$$

where  $\mu, \theta_1, \dots, \theta_q$  are constant coefficients. In the definition above,  $\varepsilon_t$  stands for the innovation at time  $t$ , and the innovations  $\varepsilon_t, \varepsilon_{t-1}, \dots$  are unobservable.

$Y_t$  defined by (54) is always stationary with  $\mathbb{E}(Y_t) = \mu$ , as the coefficient  $\theta_1, \dots, \theta_q$  do not vary over time. It is easy to show that

$$\begin{aligned} \text{var}(Y_t) &= \mathbb{E}\{(Y_t - \mu)^2\} = \mathbb{E}\{(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q})^2\} \\ &= \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2), \end{aligned} \quad (55)$$

which is independent of  $t$ . Furthermore, noticing the common white noise terms in both  $Y_{t+1}$  and  $Y_t$  are  $\varepsilon_t, \dots, \varepsilon_{t-q+1}$ , we obtain

$$\begin{aligned} \text{cov}(Y_{t+1}, Y_t) &= \mathbb{E}\{(Y_{t+1} - \mu)(Y_t - \mu)\} \\ &= \mathbb{E}\{(\varepsilon_{t+1} + \theta_1 \varepsilon_t + \dots + \theta_q \varepsilon_{t-q+1})(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_{q-1} \varepsilon_{t-q})\} \\ &= \sigma^2(\theta_1 + \theta_2 \theta_1 + \dots + \theta_q \theta_{q-1}), \end{aligned} \quad (56)$$

which is also independent of  $t$ . In general, for any  $1 \leq k \leq q$ , we have

$$\text{cov}(Y_{t+k}, Y_t) = \sigma^2(\theta_k + \theta_{k+1} \theta_1 + \dots + \theta_q \theta_{q-k}). \quad (57)$$

We can extend the above expression for  $k = 0$  by adopting the notation  $\theta_0 = 1$ . For  $k > q$ , there is no overlapping subsets of  $\{\varepsilon_t\}$ . Hence  $\text{cov}(Y_{t+k}, Y_t) = 0$ . Since the RHS of (57) is independent of  $t$ ,  $Y_t$  is always stationary. And for any MA( $q$ ) process, the ACF cuts off at  $q$ , i.e.  $\rho(k) = 0$  for any  $|k| > q$  [9].

The MA( $q$ ) process defined by (54) is said to be *invertible* if the  $q$  roots of the characteristic equation  $1 + \theta_1 x + \dots + \theta_q x^q = 0$  are outside of the unit circle [9].

In general, we may consider an MA( $\infty$ ) model

$$Y_t = \mu + \varepsilon_t + \sum_{j=1}^{\infty} \theta_j \varepsilon_{t-j}, \quad (58)$$

where the coefficients  $\theta_j$  satisfy the condition  $\sum_j \theta_j^2 < \infty$ . Under this condition,  $\mathbb{E}(Y_t) = \mu$ , and the ACVF of  $Y_t$  has the expression

$$\gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+|k|}, \quad k = 0, \pm 1, \pm 2, \dots, \quad (59)$$

where  $\theta_0 = 1$ .

The autoregressive (AR) model expresses explicitly the present value  $Y_t$  as a linear regression of its lagged values with innovation  $\varepsilon_t$  as noise:

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad (60)$$

where  $\varepsilon_t \sim WN(0, \sigma^2)$ , and  $c, \phi_1, \dots, \phi_p$  are parameters. We write  $Y_t \sim AR(p)$ . AR processes are always invertible [12]; however, not all AR processes are stationary. For example, a necessary and sufficient condition for an AR(1) process  $Y_t = c + \phi Y_{t-1} + \varepsilon_t$  to be covariance stationary is that  $|\phi| < 1$ . For a stationary AR(1) process with zero mean (i.e.  $c = 0$ ), it follows from recursive substitution that

$$Y_t = \varepsilon_t + \phi Y_{t-1} = \varepsilon_t + \phi(\varepsilon_{t-1} + \phi Y_{t-2}) = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 Y_{t-2} \quad (61)$$

$$= \varepsilon_t + \phi \varepsilon_{t-1} + \dots + \phi^k \varepsilon_{t-k} + \phi^{k+1} Y_{t-k-1} \longrightarrow \sum_{j=1}^{\infty} \phi^j \varepsilon_{t-j} \quad (62)$$

in mean square, since

$$\lim_{k \rightarrow \infty} \mathbb{E} \left\{ \left( Y_t - \sum_{j=0}^k \phi^j \varepsilon_{t-j} \right)^2 \right\} = \lim_{k \rightarrow \infty} |\phi|^{2(k+1)} \mathbb{E} (Y_{t-k-1}^2) = 0. \quad (63)$$

This shows that a stationary AR(1) process is effectively a stationary MA( $\infty$ ) process. Hence its ACF does not cut off at a finite lag, which is an important characteristic distinguishing AR from MA.

We can extend the idea above to high-order AR processes. Define the lag operator  $L$  as

$$LY_t = Y_{t-1}, \quad L^k Y_t = Y_{t-k} \quad \text{for } k = \pm 1, \pm 2, \dots$$

AR( $p$ ) model (60) can be written as

$$\Phi(L)Y_t = c + \varepsilon_t, \quad (64)$$

where the characteristic polynomial  $\Phi(\cdot)$  is defined as

$$\Phi(x) = 1 - \phi_1 x - \dots - \phi_p x^p.$$

Let  $\alpha_1^{-1}, \dots, \alpha_p^{-1}$  be the roots of the equation  $\Phi(x) = 0$ , i.e.

$$\Phi(x) = (1 - \alpha_1 x) \dots (1 - \alpha_p x).$$

Suppose  $|\alpha_j| < 1 \forall 1 \leq j \leq p$ . Then, using Taylor expansion,  $\Phi(x)^{-1}$  maybe expressed as

$$\Phi(x)^{-1} = \prod_{j=1}^p (1 - \alpha_j x)^{-1} = \prod_{j=1}^p (1 + \alpha_j x + \alpha_j^2 x^2 + \alpha_j^3 x^3 + \dots).$$

The above expression can further be expressed as

$$\Phi(x)^{-1} = 1 + \sum_{k=1}^{\infty} \theta_k x^k$$

where  $\theta'_k$ s are determined by  $\alpha_1, \dots, \alpha_p$ , and  $\sum_k |\theta_k| < \infty$  [9]. In fact

$$|\theta_k| \leq \left( \max_{1 \leq j \leq p} |\alpha_j| \right)^k, \quad k = 1, 2, \dots$$

[9] Therefore, from (64) we have

$$\begin{aligned} Y_t &= \Phi(L)^{-1}(c + \varepsilon) \\ &= c' + \Phi(L)^{-1}\varepsilon_t = c' + \varepsilon_t + \sum_{k=1}^{\infty} \theta_k \varepsilon_{t-k}, \end{aligned} \quad (65)$$

i.e.  $Y_t \sim MA(\infty)$ . Hence  $Y_t$  is stationary. The MA( $\infty$ ) representation of a stationary AR process implies that  $Y_t$  depends on  $\varepsilon$  and its lagged values, and is uncorrelated with the future white noise terms. Such a process is also called a *causal process*. By

symmetry, any invertible MA( $q$ ) is a stationary AR( $\infty$ ). For example, consider the MA(1) process  $Y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ . Solve for  $\varepsilon_t$  we get

$$\begin{aligned}\varepsilon_t &= Y_t - \theta\varepsilon_{t-1} \\ &= Y_t - \theta(Y_{t-1} - \theta\varepsilon_{t-2}) \\ &= Y_t - \theta Y_{t-1} + \theta^2(Y_{t-2} - \theta\varepsilon_{t-3}) \\ &\dots \\ &\Rightarrow Y_t = \varepsilon_t + \theta Y_{t-1} - \theta^2 Y_{t-2} + \theta^3 Y_{t-3} - \dots\end{aligned}$$

Here comes the theorem. The AR( $p$ ) process defined by (60) is stationary if the  $p$  roots of the characteristic equation  $1 - \phi_1x - \dots - \phi_p x^p = 0$  are outside of the unit circle (the modulus of all complex roots are bounded by 1). Furthermore the ACF of a stationary AR( $p$ ) process decays at an exponential rate, and the PACF cuts off at  $p$ , i.e.  $\pi(k) = 0$  for any  $k > p$  [9].

### C. Augmented Dickey-Fuller (ADF) Test

There are three variations of the ADF test, each one designed for a different alternative hypothesis: no constant and no trend, with constant but no trend, with constant and trend. With  $Z_t = \nabla Y_t = Y_t - Y_{t-1}$ ,

$$Y_t = \alpha Y_{t-1} + b_1 Z_{t-1} + \dots + b_p Z_{t-p} + \varepsilon_t, \quad (66)$$

$$Y_t = \mu + \alpha Y_{t-1} + b_1 Z_{t-1} + \dots + b_p Z_{t-p} + \varepsilon_t, \quad (67)$$

$$Y_t = \mu + \beta t + \alpha Y_{t-1} + b_1 Z_{t-1} + \dots + b_p Z_{t-p} + \varepsilon_t. \quad (68)$$

The null hypothesis is  $H_0 : \alpha = 1$  and the alternative hypothesis is  $H_1 : \alpha < 1$ . The former represents nonstationary process whereas the latter indicates stationary process.

Let  $\hat{\alpha}$  be the least squares estimator for  $\alpha$ , and  $\text{SE}(\hat{\alpha})$  denote its standard error. We reject  $H_0$  when the ADF test statistic  $W$  is smaller than a critical value, where  $W = (\hat{\alpha} - 1)/\text{SE}(\hat{\alpha})$ . Under the null hypothesis  $W$  does NOT follow  $t$ -distribution or normal distribution. Instead,  $W$  converges to a non-standard distribution which is a functional of Brownian motions, and the distributions under model (66)-(68) are different from each other. Therefore, we cannot use critical values from  $t$ -distributions. Fortunately the quantiles or critical values of those distributions have been tabulated in many places; see [16]. The implementation (in R) of ADF test we use in this project can be found below.

Suppose the series is not trending, (a) if the ADF test (without trend) rejects, then apply ARMA model directly; (b) If the ADF test (without trend) does not reject, then apply ARMA model after taking difference (maybe several times). Suppose the series is trending (a) If the ADF test (with trend) rejects, then apply ARMA model after detrending the series (b) If the ADF test (with trend) does not reject, then apply ARMA model after taking difference (maybe several times).

The choice between the above three tests can be guided by the nature of data, revealed by plotting the series. If not obvious, more than one test equation can be used to check the robustness of a test conclusion [17].

```
# aDF
aDF.test<-function (x, kind = 3, k = trunc((length(x)- 1)^(1/3)))
{

#kind = the kind of test undertaken
#kind = 1 ==> No constant no trend
#kind = 2 ==> Constant
#kind = 3 ==> Constant and trend

#the null is ALWAYS non stationarity

if (NCOL(x) > 1)
  stop("x_is_not_a_vector_or_univariate_time_series")

if (any(is.na(x)))
  stop("NAs_in_x")

if (k < 0)
  stop("k_negative")

DNAME <- deparse(substitute(x))

k <- k + 1
y <- diff(x)
n <- length(y)
z <- embed(y, k)
yt <- z[,1]
xt1 <- x[k:n]
tt <- k:n

if (kind == 1)
{
```

```





```

```

if (interpol == min(tablep))
  warning("p-value_smaller_than_printed_p-value")
else warning("p-value_greater_than_printed_p-value")

PVAL <- interpol

PARAMETER <- k - 1
METHOD <- "Augmented_Dickey-Fuller_Test"
names(STAT) <- "Dickey-Fuller"
names(PARAMETER) <- "Lag_order"

structure(list(statistic = STAT, parameter = PARAMETER, alternative =
"The_series_is_stationary",
p.value = PVAL, method = METHOD, data.name = DNAME), class = "htest")
}

```

#### D. Forecasting ARMA Models

When the time series follows an AR equation, the  $m$ -step-ahead predictor can be recursively computed based on the AR equation. The mean squared predictive error can also be calculated in a recursive manner. As an example, let us consider  $AR(1)$

$$Y_t = \phi Y_{t-1} + \varepsilon_t,$$

where  $|\phi| < 1$ ,  $\varepsilon \sim WN(0, \sigma^2)$ . Further assume that

$$\mathbb{E}(\varepsilon_t | Y_{t-1}, Y_{t-2}, \dots) = 0 \text{ for all } t. \quad (69)$$

Then the one-step ahead predictor is

$$Y_T(1) = \mathbb{E}_T\{\phi Y_T + \varepsilon_{T+1}\} = \phi Y_T. \quad (70)$$

Its MSPE is

$$MSPE\{Y_T(1)\} = \mathbb{E}[\{Y_{T+1} - Y_T(1)\}^2] = \mathbb{E}(\varepsilon_{T+1}^2) = \sigma^2. \quad (71)$$

Similarly, the two-step ahead predictor is

$$Y_T(2) = \mathbb{E}_T\{Y_{T+2}\} = \mathbb{E}_T\{\varepsilon_{T+2} + \phi Y_{T+1}\} = \phi^2 Y_T. \quad (72)$$

What is not predicted in  $Y_{T+2}$  is the future noises  $\varepsilon_{T+2} + \phi \varepsilon_{T+1}$ . Therefore

$$MSPE\{Y_T(2)\} = (1 + \phi^2)\sigma^2. \quad (73)$$

In general, for any  $k \geq 1$ , we have

$$Y_T(k) = \phi^k Y_T \quad (74)$$

and

$$MSPE\{Y_T(k)\} = (1 + \phi^2 + \dots + \phi^{2(k-1)})\sigma^2. \quad (75)$$

Hence a 95% predictive interval for  $Y_{T+k}$  is given as

$$Y_T(k) \pm 1.96\sigma\sqrt{1 + \phi^2 + \dots + \phi^{2(k-1)}}. \quad (76)$$

When the forecast horizon  $k$  increases, the MSPE increases. When  $k \rightarrow \infty$ ,  $Y_T(k) \xrightarrow{P} 0$ , the expected value of the time series, and by Taylor expansion,

$$MSPE\{Y_T(k)\} \rightarrow \frac{\sigma^2}{1 - \phi^2} = \text{var}(Y_{T+k}). \quad (77)$$

This indicates that the long-term forecasting is difficult, as the best prediction is the mean of the time series. In fact, the prediction error approaches the unconditional variance  $\text{var}(Y_t)$ .

For a causal and invertible  $AR(p, q)$  model

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (78)$$

where  $\varepsilon_t \sim WN(0, \sigma^2)$  satisfying the condition (69). An added complication here is to recover unobserved innovations  $\{\varepsilon_t, t \leq T\}$  from observations  $\{Y_t, t \leq T\}$ . The invertibility implies that  $Y_t$  has the following  $AR(\infty)$  representation

$$Y_t - \mu = \varepsilon + \sum_{\ell=1}^{\infty} \psi_{\ell}(Y_{t-\ell} - \mu), \quad (79)$$

where  $\mu = c/(1 - \phi_1 - \dots - \phi_p)$ , and  $\psi_{\ell}$  are determined by  $\theta_i$  and  $\phi_j$  in the original model (78). This indicates that we can recover the innovations  $\{\varepsilon_t, t \leq T\}$  from observations  $\{Y_t, t \leq T\}$  [9]. Hence

$$\varepsilon_T(k) \equiv \mathbb{E}_T(\varepsilon_{T+k}) = \begin{cases} 0 & \text{if } k \geq 1 \\ \varepsilon_{T+k} & \text{if } k \leq 0 \end{cases}, \quad (80)$$

namely, the best prediction for the future innovation is 0 and for the realized innovation is itself. (Note, some authors use  $\{e_t, t \leq T\}$  to denote observed residual/innovation; see section 8.8 [23] for example.) Therefore, we can derive the one-step ahead predictor for the process (78):

$$Y_T(1) = c + \phi_1 Y_T + \cdots + \phi_p Y_{T-p+1} + \theta_1 \varepsilon_{T-1} + \cdots + \theta_q \varepsilon_{T-q+1}. \quad (81)$$

In general, for any  $k \geq 1$ ,

$$Y_T(k) = c + \phi_1 Y_T(k-1) + \cdots + \phi_p Y_T(k-p) + \theta_1 \varepsilon_T(k-1) + \cdots + \theta_q \varepsilon_T(k-q), \quad (82)$$

where  $\varepsilon_T(k)$  is defined as in (80), and  $Y_T(j) = Y_{T+j}$  for all  $j \leq 0$ .

It is easy to see that

$$\text{MSPE}\{Y_T(1)\} = \text{var}(\varepsilon_{T+1}) = \sigma^2. \quad (83)$$

To compute  $\text{MSPE}\{Y_T(k)\}$  for  $k > 1$ , we appeal to the causal property of model (78). We can show that

$$\text{MSPE}\{Y_T(k)\} = \mathbb{E}[\{Y_{T+k} - Y_T(k)\}^2] = \sigma^2 \left\{ 1 + \sum_{\ell=1}^{k-1} \psi_{\ell}^2 \right\}, \quad (84)$$

which increases as  $k$  increases, indicating that the prediction error increases when the forecast horizon increases. Similar to AR, long-term forecasting is difficult. When  $k \rightarrow \infty$ , we have  $Y_T(k) \xrightarrow{P} \mu$ . In other words, the best predictor is the unconditional mean of the time series itself. Also,  $\text{MSPE}\{Y_T(k)\} \rightarrow \text{var}(Y_{T+k})$ , same as the unpredicted one.

The techniques presented are practically applicable to ARIMA models for which the parameters can be replaced by their estimators obtained from the differenced observations.

### E. Damped Trend Methods

In conjunction with the smoothing parameters  $\alpha$  and  $\beta$ , this method also includes a damping parameter  $0 \leq \phi \leq 1$ :

$$\text{Forecast equation } \hat{Y}_{t+k|t} = \ell_t + (\phi + \phi^2 + \cdots + \phi^k) F_t \quad (85)$$

$$\text{Smoothing equation } \ell_t = \alpha Y_t + (1 - \alpha)(\ell_{t-1} + \phi F_{t-1}) \quad (86)$$

$$\text{Trend equation } F_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi F_{t-1}, \quad (87)$$

Thus, the growth for the one-step forecast  $\hat{Y}_{T+1|T}$  is  $\phi F_T$ , and the growth is damped by a factor of  $\phi$  for each additional future time period. If  $\phi = 1$ , this method gives the same forecasts as Holt's linear method. For  $0 < \phi < 1$ , as  $h \rightarrow \infty$  the forecasts approach an asymptote given by  $\ell_t + \phi F_t / (1 - \phi)$ . This means that short-run forecasts are trended while long-run forecasts are constant. We usually restrict  $\phi > 0$  to avoid a negative coefficient being applied to  $F_{t-1}$  in (87), and  $\phi \leq 1$  to avoid  $F_t$  increasing exponentially [20]. In practice, however,  $\phi$  is rarely less than 0.8 as the damping has a very strong effect for smaller values. Values of  $\phi$  close to 1 will mean that a damped model is not able to be distinguished from a non-damped model. For these reasons, we usually restrict  $0.8 \leq \phi \leq 0.98$  [23].

### F. More on Nonstationary variables

Nonstationary variables that wander up and down, trending in one direction and then the other, are said to be possess a *stochastic trend*. Definite trends, upward or downward, can be attributable to a stochastic trend or a *deterministic trend*, or sometimes both. Variables that are stationary after removing a deterministic trend are called *trend stationary*.

Consider an AR(1) model fluctuating around a deterministic linear trend  $(\mu + \delta t)$ . In this case, we can let the "detrended" series  $(Y_t - \mu - \delta t)$  behave like an autoregressive model

$$(Y_t - \mu - \delta t) = \phi[Y_{t-1} - \mu - \delta(t-1)] + \nu_t, \quad |\phi| < 1. \quad (88)$$

Equation (88) is an example of a trend stationary process. The detrended series  $(Y_t - \mu - \delta t)$  has a constant variance, and covariance that depend only on the time separating observations, not the time at which they are observed. In other words, the detrended series is stationary;  $Y_t$  is stationary around the deterministic trend.

Now consider a *random walk* without drift defined as

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (89)$$

where  $\varepsilon_t \sim \text{WN}(0, \sigma^2)$ . A random walk is an AR(1) model with the AR coefficient  $\phi$  equal to 1, whose characteristic equation  $\Phi(x) = 1 - x$  has a unit root, i.e. a root on the unit circle. A random walk appears to wander slowly upward and downward with no real pattern; the sample means calculated from subsamples of observations will be dependent on the sample period, a characteristic of nonstationary series. To verify this, we may assume that the process starts at an arbitrary point  $Y_0 = c$  and rewrite  $Y_t$  as  $Y_t = c + \varepsilon_1 + \cdots + \varepsilon_t$ , then

$$\text{var}(Y_t) = \text{var}(Y_{t-1} + \varepsilon_t) = \text{var}(c + \varepsilon_1 + \cdots + \varepsilon_t) = t\sigma^2,$$

which increases linearly with time  $t$ . The component  $\sum_{k=1}^t \varepsilon_k$  is called the stochastic trend. This term arises because a stochastic component  $\varepsilon_t$  is added for each time  $t$ , and because it causes the time series to trend in unpredictable directions. We can extend the random walk model by adding a constant term to (89):

$$Y_t = \delta + Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots. \quad (90)$$

This model is known as the *random walk with drift*. We can extend the random walk even further by adding a time trend:

$$Y_t = \alpha + \delta t + Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots \quad (91)$$

It is easy to show that neither (90) nor (91) are stationary.

We noted that regressions involving variables with a deterministic trend, and no stochastic trend, did not present any difficulties providing the trend was included in the regression relationship, or the variable were detrended [17]. Now we consider the implication of estimating regressions involving variables with stochastic trends. From now on, since stochastic trends are the most prevalent source of nonstationarity [17], and they introduce special problems, when we refer to nonstationary variables, we generally mean variables that are neither stationary nor trend stationary.

### G. Multivariate time series analysis

Let us consider a general *dynamic regression model* that contains two time series  $Y_t$  and  $X_t$  and their lagged terms:

$$Y_t = c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_0 X_t + \theta_1 X_{t-1} + \dots + \theta_q X_{t-q} + \varepsilon_t. \quad (92)$$

Note that this model is more general than the ARMA( $p, q$ ) model and is referred to as an *autoregressive distributed lag model*, abbreviated as an ARDL( $p, q$ ) model. The AR component of the name ARDL comes from the regression of  $Y_t$  on lagged values of itself; the DL component comes from the distributed lag effect of the lagged explanatory variable  $X_t$ .

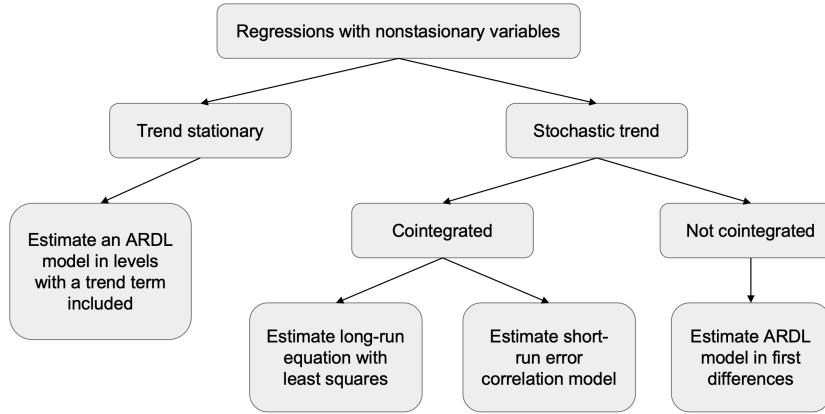
A consequence of proceeding with regression involving nonstationary variables with stochastic trends is that OLS estimates no longer have approximate normal distribution in large sample. That means precision of estimation may not be what it seems to be and conclusions about relationships between variables could be wrong. One particular bad consequence is that two totally independent random walks can appear to have a strong linear relationship when none exists [17]. This kind of regressions have been given the name *spurious regression*.

As a general rule, to avoid the problem of spurious regression, nonstationary time-series variables should not be used in regression models. However, there is an exception to this rule. If two series  $Y_t$  and  $X_t$  are nonstationary I(1) variables, then we expect their difference, or any linear combination of  $Y_t$  and  $X_t$ , such as  $e_t = Y_t - \beta_1 - \beta_2 X_t$ , to be I(1) as well. However, if  $e_t$  is a stationary process, then  $Y_t$  and  $X_t$  are said to be *cointegrated* [17]. The term *cointegration* was introduced by Granger [14]. A formal definition is given as follow [9]: a vector time series  $\mathbf{Y}_t$  is said to be cointegrated with order  $(k, h)$  ( $k \geq h \geq 1$ ), denoted as  $\mathbf{Y}_t \sim \text{CI}(k, h)$ , if

- i) all component series of  $Y_t$  are I( $k$ ), and,
- ii) there exists a nonzero vector  $\beta$  such that  $\beta^\top \mathbf{Y}_t \sim \text{I}(k-h)$ .

The most frequently used cointegration model is CI(1, 1), i.e. all the components of  $\mathbf{Y}_t$  are I(1) and  $\beta^\top \mathbf{Y}_t$  is stationary.

Going back to our example of  $Y_t$  and  $X_t$ , a natural way to test whether  $Y_t$  and  $X_t$  are cointegrated is to test the stationarity of the OLS residuals,  $\hat{e}_t = Y_t - \beta_1 - \beta_2 X_t$  using an ADF test. If these two variables are I(1) and cointegrated, we can estimate a regression relationship by either estimating a least-squares equation (long-run relation) between the I(1) them or by estimating a nonlinear least-squares *error correction model* (ECM) which embeds the I(1) variables [7]. If the variables are I(1) and not cointegrated, we need to estimate a relationship in their first differences, with or without the constant term. These options are summarized in Figure 15.



**Figure 15:** Strategies of dealing with nonstationary time series variables

### H. Data Sets and Code

You can reproduce our visualizations and modeling results by downloading and walking through the code on our [GitHub](#) yourself.

## REFERENCES

- [1] Hirotugu Akaike. Information theory and an extension of the maximum likelihood principle. 1973.
- [2] CSSE at Johns Hopkins University. Covid-19 data repository. Data retrieved from <https://github.com/CSSEGISandData/COVID-19>.
- [3] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] Robert G. Brown. *Statistical Forecasting for Inventory Control*. McGraw-Hill, 1959.
- [5] Chung Chen and Lon-Mu Liu. Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88(421):284–297, 1993.
- [6] Catherine Cote. 5 principles of data ethics for business. *Harvard Business School*, 2021. From <https://online.hbs.edu/blog/post/data-ethics>.
- [7] Robert F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987.
- [8] Jianqing Fan and Qiwei Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, 2003.
- [9] Jianqing Fan and Qiwei Yao. *The Elements of Financial Econometrics*. Cambridge University Press, 2017.
- [10] Yahoo Finance. Nyse arca airline index. Data retrieved from <https://finance.yahoo.com/quote/%5EXAL?p=%5EXAL>.
- [11] Everette S. Gardner and Ed. Mckenzie. Forecasting trends in time series. *Management Science*, 31(10):1237–1246, 1985.
- [12] Gloria González-Rivera. *Forecasting for Economics and Business*. Routledge, 2013.
- [13] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.
- [14] C. W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130, 1981.
- [15] Jeff Hallman. *tis: Time Indexes and Time Indexed Series*, 2020. R package version 1.38.
- [16] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- [17] C. Hill, W. Griffiths, and G. Lim. *Principles of Econometrics, 5th Edition*. Wiley, 2018.
- [18] Charles C. Holt. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10, 2004.
- [19] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, and F. Yasmeen. *forecast: Forecasting Functions for Time Series and Linear Models*, 2020. R package version 8.13.
- [20] R. Hyndman, A. Koehler, K. Ord, and R. Snyder. *Forecasting with Exponential Smoothing The State Space Approach*. Springer, 2008.
- [21] Rob Hyndman. Comparing HoltWinters() and ets(), 2011.
- [22] Rob Hyndman. *fpp2: Data for "Forecasting: Principles and Practice"* (2nd Edition), 2020. R package version 2.4.
- [23] Rob Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice* (2nd ed). OTexts, 2018.
- [24] Christian Kleiber and Achim Zeileis. *Applied Econometrics with R*. Springer-Verlag, New York, 2008. ISBN 978-0-387-77316-2.
- [25] United States Department of Transportation. Bureau of transportation statistics t-100 market data. Data retrieved from BTS TranStats, [https://www.transtats.bts.gov/Data\\_Elements.aspx?Data=2](https://www.transtats.bts.gov/Data_Elements.aspx?Data=2).
- [26] United States Department of Transportation. Bureau of transportation statistics t-100 market data. Data retrieved from FRED, Federal Reserve Bank of St. Louis, <https://fred.stlouisfed.org/series/RPM>.
- [27] B. Pfaff. *Analysis of Integrated and Cointegrated Time Series with R*. Springer, New York, second edition, 2008. ISBN 0-387-27960-1.
- [28] Bernhard Pfaff. *VAR, SVAR and SVEC Models: Implementation Within R Package vars*, 2008. *Journal of Statistical Software* 27(4).
- [29] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.
- [30] Jeffrey A. Ryan and Joshua M. Ulrich. *xts: eXtensible Time Series*, 2020. R package version 0.12.1.
- [31] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461 – 464, 1978.
- [32] Ritei Shibata. Asymptotically Efficient Selection of the Order of the Model for Estimating Parameters of a Linear Process. *The Annals of Statistics*, 8(1):147 – 164, 1980.
- [33] M. Strohmeier, X. Olive, J. Lübbe, M. Schäfer, and V. Lenders. Crowdsourced air traffic data from the opensky network 2019–2020 (version v21.12), 2021. DOI:10.5281/zenodo.5815402 (Accessed February 1, 2022).
- [34] G. C. Tiao and G. E. P. Box. Modeling multiple time series with applications. *Journal of the American Statistical Association*, 76(376):802–816, 1981.
- [35] Adrian Trapletti and Kurt Hornik. *tseries: Time Series Analysis and Computational Finance*, 2020. R package version 0.10-48.
- [36] Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2018. R package version 0.7.6.
- [37] Peter R. Winters. Forecasting sales by exponentially weighted moving averages. *Management Science*, 6(3):324–342, 1960.