Stage 1: Brief Background and Define Goal

A revenue passenger mile (RPM) is a metric used in the transportation industry, specifically the airline industry, that displays the distance (in miles) travelled by paying customers (Kenton, 2022). It is basically measuring the demand for air transport, whereby an increase in RMP means more passengers are using air transportation services (Cederholm, 2014). This metric is influenced by changes in the national economy, whereby as the national GDP increases, the demand for airline travel increases (National Academy of Sciences, 2021).

The COVID-19 pandemic posed unprecedented challenges for the global and U.S. airline industry, sharply contrasting with the previous upward trend in RPM. Government-imposed restrictions and health concerns resulted in a dramatic decline in passenger confidence and international travel demand. Airlines responded with flight cancellations, employee furloughs, and fleet groundings, leading to a significant and prolonged drop in RPM.

The goal of this research assignment is to utilise forecasting methods to assess air traffic statistics for the United States and predict the effects of the pandemic on the aviation industry in the U.S. This forecasting modelling aims to provide insights into future outcomes, allowing for better planning, decision-making, and resource allocation.

Stage 2: Time Series Exploration and Visualisation

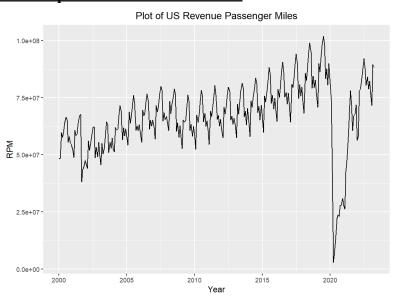


Figure 2.1: Dataset of US Revenue Passenger Miles from January 2000 to April 2023

Figure 2.1 reveals a distinct seasonality pattern accompanied by a cyclical nature in the data representing Revenue Passenger Miles (RPM) in the United States. This cyclical pattern implies that RPM experiences regular fluctuations over time, with a strong seasonality pattern. Notably, amidst these fluctuations, we can observe from **Figure 2.1** whereby it displays an upward trend, indicating a consistent increase in the demand for airline transportation as the years progress. This upward trajectory suggests that, on the whole, more people are opting for air travel over time.

However, the RPM data is far from immune to external factors, as highlighted in **Figure 2.2.** Since the turn of the century, the U.S. economy faced two significant recessions - one occurring between March and November 2001 and another during the infamous Great Recession spanning December 2007 to June 2009 (National Academy of Sciences, 2021). Both of these economic downturns exerted substantial influence on the demand for aviation travel. The 2001 recession, which was notably brief, can be attributed to the Y2K Scare and the tragic events of the 9/11 attacks. (Amadeo, 2021). This event is reflected in the slightly different seasonality pattern in the initial two years. Notably, **Figure 2.2** underscores this adverse impact by showing a substantial drop in RPM from 67 million in August 2001 to a mere 38 million in September 2001, reflecting the heightened safety concerns and reluctance to fly following the 9/11 attacks.

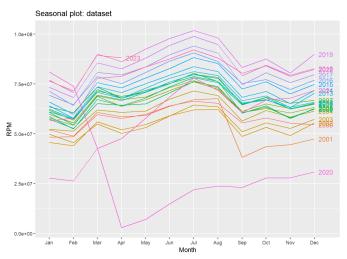


Figure 2.2: Seasonal plot of the US RPM dataset from January 2000 to April 2023.

Since the terrorist attack, the USA has significantly increased airport security screening to ensure the safety of its national welfare, which, as noted by Blalock et al. (2007), reduced the demand for air travel. This heightened focus on security also contributed to the fear among air travellers, causing a negative impact on the number of passengers. It was estimated that the airline industry incurred substantial losses, with approximately \$1.1 billion in revenue lost due to the reduction in demand. This amount represents a significant 11 percent of the total losses estimated by the General Accounting Office for the airline industry as a consequence of the 9/11 attacks.

2002 to 2007:

The period from mid-2002 to 2007 in the airline industry signifies a phase of recovery and growth after the disruptive aftermath of the 9/11 attacks and subsequent economic downturn. This could be observed in the rising trend from 2002 till a gradual decline in 2007 as the citizens' confidence to travel returned and security measures improved. During this time, a relatively stable U.S. economy with consistent growth provided passengers with more disposable income, contributing to increased air travel. Airlines adapted by implementing cost-cutting measures and introducing new services, further fueling industry growth. Additionally, the resurgence of business travel, often a significant factor in RPM trends, played a role as the economy improved.

2008 to 2009:

The period from 2008 onwards in the airline industry was characterised by the global financial crisis (The Great Recession) of 2008 and volatile oil prices, prompting airlines to implement cost-saving measures, including capacity reductions and mergers. These changes reshaped the industry's landscape as airlines aimed to optimise operations and reduce costs. The industry faced challenges due to reduced consumer spending, job losses, and declining travel during the Great Recession, which had a direct impact on RPM. Technology and online booking platforms continued to grow, enhancing the passenger experience and streamlining operations. Despite these challenges, the industry gradually recovered as the economy stabilised, and airlines adapted by focusing on improving services, enhancing fuel efficiency, and expanding routes to meet changing passenger expectations, demonstrating resilience and adaptability in a changing landscape.

2010 onwards:

Additionally, from **Figure 2.1** we can also see the downward trend in the RPM from 2007 to 2009 due to the Great Recession affecting people's spending power before started to increase again in 2010. From 2010 until 2020, we can see a steady upward trend with a strong seasonality pattern. However, up until 2020, the pattern was cut off by a structural break whereby the RPM started to plummet drastically, being the lowest compared to the other years since 2020. This is caused by the recent COVID-19 pandemic that started around December 2019 being the peak in 2020 and has raised a general fear of catching and spreading COVID-19 through air travel. Therefore, most countries are forced to impose international flight restrictions and stay-at-home orders, which has significantly impacted the aerospace and aviation industries that are very dependent on commercial passenger travel. Additionally, on March 19, the US State Department issued a worldwide Level 4: Do Not Travel Advisory (O'Hare & Hardingham-Gill, 2020), which is the main reason why the RPM plummeted drastically from 73 million in February to 2.9 million in April (see **Figure 2.2**), being the lowest ever recorded since 2000. Afterwards, the RPM increased gradually but it took three years to be able to recover to the 2019 level.

Stage 3: Pre-Process Data

To obtain a better and more accurate forecasting result, it is necessary to pre-process the data so as to only include a time frame that contains fewer or no structural breaks. In our case, the 2001 Recession and the Great Recession from 2007 to 2009 significantly affected the reliability of the dataset. Hence, for our "full dataset", we will exclude the time frame that includes the Recession periods and only include the time frame after the Great Recession which starts from Jan 2010 until April 2023. Hence, if we plot our "full dataset" as shown in **Figure 3.1**, we can see that there is no cyclical pattern up until 2020, which enables our forecasting model to fit efficiently.

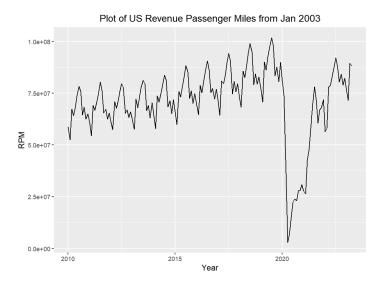


Figure 3.1: Plot of US Revenue Passenger Miles after Pre-Processing

Stage 4: Partition Series

We will partition the full dataset (January 2010 to April 2023) into pre-Covid and Covid datasets. **The pre-Covid** dataset includes data prior to the pandemic which is from January 2010 until January 2020 and will be split into a training set and test set on a 75-25 ratio (**Figure 4.1**). The **train set** (75%) contains the beginning 97 observations, from January 2010 to January 2018. While the **test set** (25%) contains the last 24 observations from February 2018 to January 2020. **Covid** dataset includes data during the COVID-19 period in the US which is from February 2020 to July 2021

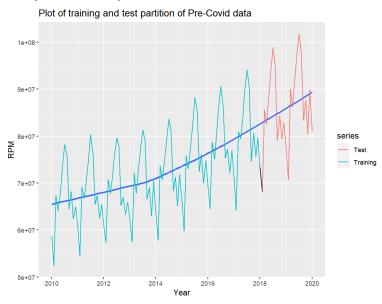


Figure 4.1: Visualisation plot of training and test partition of Pre-Covid data

Stage 5: Apply Forecasting Methods/Models

In this stage, we will apply one out of four simple forecasting methods (Drift, Mean, Naive, or Seasonal Naive) and two ETS exponential smoothing models into the Pre-Covid dataset. Additionally, we will use ets() function in R to automatically compute an ETS model, resulting in four forecasting methods/models.

Simple Forecasting Method

Among the four simple forecasting methods, it is expected that the most logical and suitable method to choose will be the seasonal naive method as the pre-covid dataset used is a monthly dataset that contains visible seasonality patterns.

Method 1 – Seasonal Naive Method

To justify this selection, we will view the plot visualisation comparing forecasts from all four methods and the accuracy measures. In **Figure 5.1**, we can see that the seasonal naive method fits the seasonality pattern of the actual data similarly compared to the other three methods. Additionally, we will choose the method that shows the smallest value of test set accuracy values in RMSE, MAPE, and MASE. Based on **Table 5.1**, we can see that the **seasonal naive method has the smallest accuracy value** as compared to the other methods. This shows that the forecasting method has the smallest forecasting error and hence is the most accurate to use.

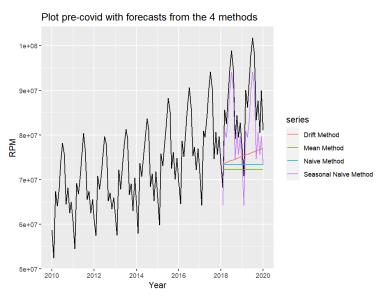


Figure 5.1: Visualisation Plot of Forecasts from Four Simple Forecasting Methods

Table 5.1: Accuracy values extracted from R for each method

| Method | RMSE | MAE | MPE | MAPE | MASE |
|----------------|------------|------------|------------|-----------|----------|
| Mean | 16,515,340 | 14,628,945 | 15.541938 | 16.219332 | 7.093669 |
| Naive | 15,611,844 | 13,741,595 | 14.2978480 | 15.230724 | 6.663387 |
| Seasonal Naive | 6,302,346 | 5,927,894 | 6.828345 | 6.828345 | 2.874474 |
| Drift | 13,967,389 | 12,006,511 | 12.0875645 | 13.273524 | 5.822034 |

However, the simple forecasting method may not be accurate enough as it does not show trend and seasonality together, like for instance the Snaive method only shows seasonality pattern with no trend. Hence, ETS Exponential Smoothing Models will be proposed in this report as it could show error terms, trends, and seasonality in one ETS model.

ETS Exponential Smoothing Models

For our two selected ETS models, since as shown in Figure 3.1, the Pre-Covid dataset shows a linear trend with a constant seasonality pattern, we will choose the models that contain additive trend and additive seasonality (Additive Holt Winters' model) with the first model using additive error term and the second model using multiplicative error term. For the automatically generated ETS model by R, the ETS model with the additive damped trend, multiplicative seasonality, and multiplicative error term is selected by minimising the AIC.

Table 5.2: Parameters Estimates for ETS Models

| | ETS [M, A, A] | ETS[A, A, A] | ETS[M, Ad, M] | | | |
|--|---|---|---|--|--|--|
| | Smoothing parameters | | | | | |
| α (level) | 0.3276 | 0.3637 | 0.342 | | | |
| β (trend) | 1e-04 = 0.0001 | 1e-04 = 0.0001 | 0.0227 | | | |
| γ (seasonal) | 1e-04 = 0.0001 | 1e-04 = 0.0001 | 1e-04 = 0.0001 | | | |
| ф (damped) | - | - | 0.98 | | | |
| Initial states | | | | | | |
| lt (level) | 65558774.335 | 65558774.5196 | 65374879.02 | | | |
| bt (slope) | 129956.0095 | 129956.5901 | 161823.6087 | | | |
| s0, s-1, s-2, s-3, s-4, s-5, s-6, s-7, s-8, s-9, s-10 | -2036876, -6346191, -827069, -3250154, 8841030, 12123255, 7995397, 3176127, -1194117, 1823829, -13011882, -7293348 | -2036876, -6346190, -827068.1, -3250154, 8841029, 12123255, 7995396, 3176127, -1194117, 1823828, -13030997, -7274233 | 0.9721, 0.9144, 0.9905, 0.9509, 1.1244, 1.1686, 1.1094, 1.0421, 0.9823, 1.0259, 0.8198, 0.8994 | | | |

Model 2 – ETS[M, A, A]: Additive Holt Winters' method with additive errors

Combined with the parameters shown in Table 5.2, the model can be written as below:

$$\begin{split} y_t &= (l_{t-1} + b_{t-1} + s_{t-m}) \, (1 + \varepsilon_t) \\ l_t &= l_{t-1} + b_{t-1} + \alpha \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t = l_{t-1} + b_{t-1} + 0.3276 \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t \\ b_t &= b_{t-1} + \beta \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t = b_{t-1} + 0.0001 \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t \\ s_t &= s_{t-m} + \gamma \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t = s_{t-m} + 0.0001 \, (l_{t-1} + b_{t-1} + s_{t-m}) \, \varepsilon_t \end{split}$$

The α value is 0.3276, meaning that 32% of the forecast is based on the most recent observations and 68% is based on the previous observations. The slope and seasonal component values are very small (0.0001), hence both these components change very little over time.

Model 3 – ETS[A, A, A]: Additive Holt Winters' method with multiplicative errors

Combined with the parameters shown in Table 5.2, the model can be written as below:

$$\begin{split} y_t &= l_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha \varepsilon_t = l_{t-1} + b_{t-1} + 0.3637 \varepsilon_t \\ b_t &= b_{t-1} + \beta \varepsilon_t = b_{t-1} + 0.0001 \varepsilon_t \\ s_t &= s_{t-m} + \gamma \varepsilon_t = s_{t-m} + 0.0001 \varepsilon_t \end{split}$$

The α value is 0.3637, meaning that 36% of the forecast is based on the most recent observations and 64% is based on the previous observations. The slope and seasonal component values are very small (0.0001), hence both these components change very little over time.

Model 4 – ETS[M, Ad, M]

Combined with the parameters shown in Table 5.2, the model can be written as below:

$$\begin{split} y_t &= (l_{t-1} + \varphi b_{t-1}) \, s_{t-m} (1 + \varepsilon_t) = (l_{t-1} + 0.98 b_{t-1}) \, s_{t-m} (1 + \varepsilon_t) \\ l_t &= (l_{t-1} + \varphi b_{t-1}) (1 + \alpha \varepsilon_t) = (l_{t-1} + 0.98 b_{t-1}) (1 + 0.342 \varepsilon_t) \\ b_t &= \varphi b_{t-1} + \beta \, (l_{t-1} + \varphi b_{t-1}) \, \varepsilon_t = 0.98 b_{t-1} + 0.0227 \, (l_{t-1} + 0.98 b_{t-1}) \, \varepsilon_t \\ s_t &= s_{t-m} + (1 + \gamma \, \varepsilon_t) = s_{t-m} + (1 + 0.0001 \, \varepsilon_t) \end{split}$$

The α value is 0.342, meaning that 34% of the forecast is based on the most recent observations and 66% is based on the previous observations. Only 2% of the slope is based on the most current slope, so a high majority of 98% is based on the most previous slope. The seasonal component value changes very little over time since its value is very small (0.0001). Additionally, the damping parameter (ϕ) is 0.98 which is almost 1, meaning that the damping is low.

Residual Diagnostic

For a reliable forecasting method/model, it is essential to have white noise since such a method/model effectively captures the changes in the data as the residuals are not independent of each other. Hence, we will use checkresiduals() function in R to determine whether the model residuals are white noise.

Seasonal Naive Method

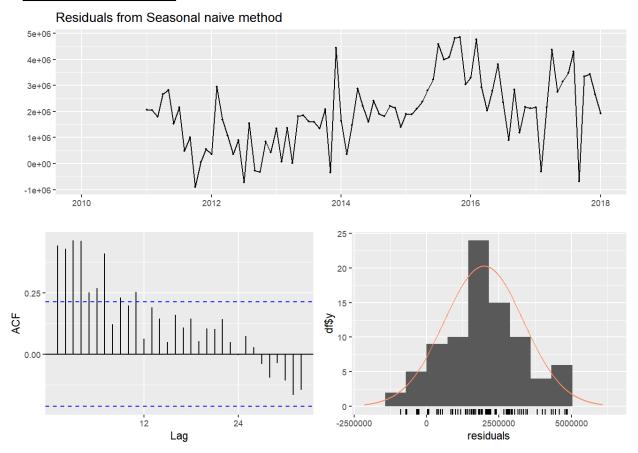


Figure 5.2: The Residual Diagnostics for ETS[M, A, A] Model

Step 1:

$$H_0 = \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$$

 H_1 : at least one of $\rho_i \neq 0$ for $i = 1, 2,, 19$

Step 2:

Significance level (α) = 0.05

Step 3:

p-value = <2.2e-16

Step 4:

Decision Rule:

Since p-value (< 2.2e-16) is a lot smaller than the significance level (0.05), we reject the null hypothesis (H_0). Thus, this model is a not white noise model since there is an autocorrelation whereby most of the lags (>95%) are outside the critical value. In conclusion, ETS[M, A, A] is a reliable forecasting model.

ETS[M, A, A]

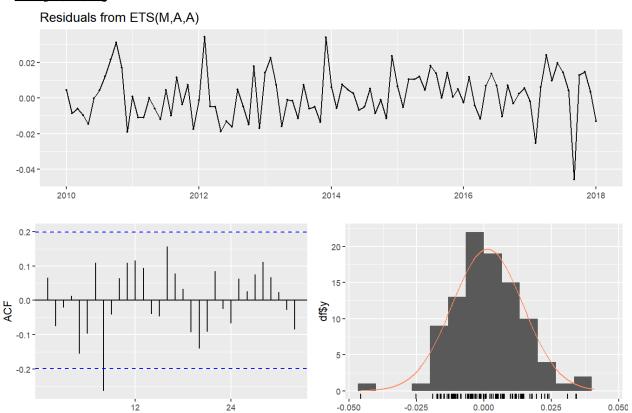


Figure 5.3: The Residual Diagnostics for ETS[M, A, A] Model

residuals

Step 1:

$$H_0 = \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$$

 H_1 : at least one of $\rho_i \neq 0$ for $i = 1, 2,, 19$

Lag

Step 2:

Significance level (α) = 0.05

Step 3:

p-value = 0.2328

Step 4:

Decision Rule:

Since p-value (0.2328) is bigger than the significance level (0.05), we do not reject the null hypothesis (H_0) . Thus, this model is a white noise model since there is no autocorrelation although there is one lag that goes beyond the blue line critical value. In conclusion, ETS[M, A, A] is a reliable forecasting model.

ETS[A, A, A]

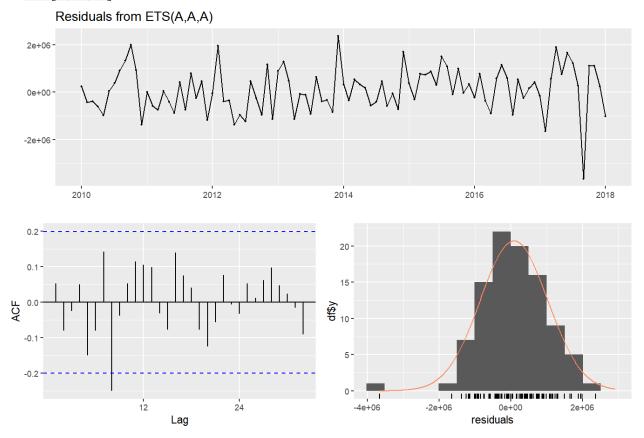


Figure 5.4: The Residual Diagnostics for ETS[A, A, A] Model

Step 1:

$$\begin{split} &H_0 = \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0 \\ &H_1 \text{: at least one of } \rho_i \neq 0 \text{ for } i = 1, \ 2, \ \ , \ 19 \end{split}$$

Step 2:

Significance level (α) = 0.05

Step 3:

p-value = 0.2781

Step 4:

Decision Rule:

Since p-value (0.2781) is bigger than the significance level (0.05), we do not reject the null hypothesis (H_0) . Thus, this model is a white noise model since there is no autocorrelation although there is one lag that goes beyond the blue line critical value. In conclusion, ETS[A, A, A] is a reliable forecasting model.

ETS[M, Ad, M]

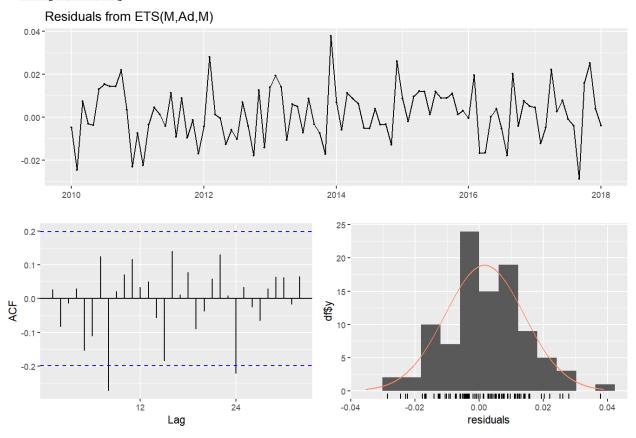


Figure 5.5: The Residual Diagnostics for ETS[M, Ad, M] Model

Step 1:

$$H_0 = \rho_1 = \rho_2 = \rho_3 = \rho_4 = \rho_5 = 0$$

 H_1 : at least one of $\rho_i \neq 0$ for $i = 1, 2,, 19$

Step 2:

Significance level (α) = 0.05

Step 3:

p-value = 0.1452

Step 4:

Decision Rule:

Since p-value (0.1452) is bigger than the significance level (0.05), we do not reject the null hypothesis (H_0) . Thus, this model is a white noise model since there is no autocorrelation although there are two lags that go beyond the blue line critical value. It is still accepted as two lags are still within 95%. In conclusion, ETS[M, Ad, M] is a reliable forecasting model.

Best Goodness of Fit

We can evaluate the best goodness of fit by looking at each of the method/models' fitted values plots and accuracy measures.

Fitted value plots

We will evaluate whether each of the methods/models underfits, fits nicely, or overfits. A good forecasting model should fit nicely with the actual data.

Plot Training Set with the Fitted Values

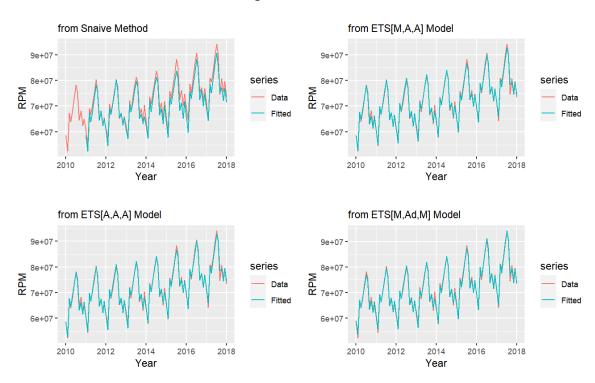


Figure 5.6: The Fitted Models for the Proposed Four Forecasting Models

Firstly, we notice that the first 12 observations in the Snaive method contain missing values. This is due to the fact that the seasonal naive method uses the latest observed value from the same season to forecast the following period. Hence, since before 2010, there are no any observations, it could not generate the forecast for 2010 and the forecast starts in 2011 using 2010 values. Moreover, we can see that the majority of the forecast fitted values (except 2012) underfit the actual data shown by how the blue line lies below the red line. This means that this model is not a good forecasting model.

The fitted values for both ETS[M, A, A] and ETS[A, A, A] seem so similar to each other, whereby the values for the first five years fit quite nicely with the actual data, but the next three years underfit a little. Contrarily, for ETS[M, Ad, M] the last three years' fitted value fit nicely, but underfit for the first two years. Hence, we can conclude that these three models are good forecasting models.

Accuracy measures (RMSE, MAPE, MAE)

RMSE calculates the root mean square error of the distance between observed and predicted values, while MAPE and MAE calculate the absolute and percentage difference between observed and predicted values. These three measures will be used to evaluate the accuracy of each method/model by choosing the lowest values. Hence, as shown in **Table 5.3**, the ETS[M, Ad, M] model has the smallest values for all three measures.

| T-11. 5 2. | T : : | 4 | |
|------------|----------|--------------|----------|
| Table 5.3: | Training | set accuracy | measures |

| | RMSE | MAPE | MAE |
|----------|-----------|-----------|-----------|
| Snaive | 2,404,002 | 2.773689 | 2,062,254 |
| M, A, A | 940,667 | 1.036011 | 743,357.9 |
| A, A, A | 939,811.4 | 1.032459 | 740,124.3 |
| M, Ad, M | 867,529.6 | 0.9872375 | 700,236 |

Produce forecast for the test set period

We will evaluate whether the forecasts of each method/model under-forecast or over-forecast the training set data. As shown in Figure 5.7, all methods/models lie below the actual data, meaning that all under-forecast. The Snaive method under-forecast the most as it does not follow the trend like the actual data. The ETS[M, A, A] and ETS[A, A, A] show a similar pattern and is located right above Snaive, while ETS[M, Ad, M] are the most similar with the actual data. Hence, in accordance with the fitted value plot and accuracy measures, ETS[M, Ad, M] is the best forecasting model.

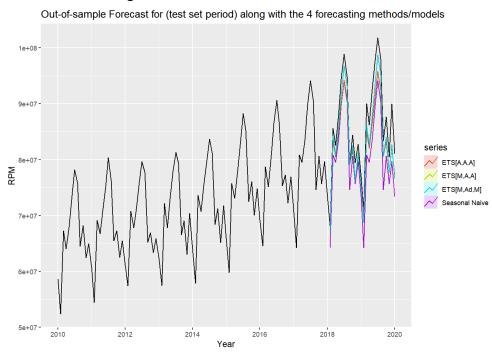


Figure 5.7: The Out-of-Sample Forecasts for the Proposed Four Forecasting Models

Stage 6: Evaluation and Comparing Forecasting Performance

Traditional Approach

We can see that ETS[M, Ad, M] also has the smallest value of RMSE, MAE, MPE, MAPE, and MASE for out of sample forecast. Hence, not only in-sample forecast, this model also fits the best with the actual data when conducting out-of-sample forecast.

Table 6.1: Out-of-sample forecast accuracy shown by the test set measures

| Method | RMSE | MAE | MPE | MAPE | MASE |
|----------|-------------|-------------|------------|-----------|-----------|
| Snaive | 6,302,346 | 5,927,894 | 6.828345 | 6.828345 | 2.874474 |
| M, A, A | 4,034,916 | 3,534,955.5 | 3.9171012 | 3.941686 | 1.714123 |
| A, A, A | 4,054,613.8 | 3,554,926.7 | 3.94472337 | 3.964252 | 1.723807 |
| M, Ad, M | 3,126,164.5 | 2,761,614 | 3.1599654 | 3.1599654 | 1.3391243 |

Time Series Crossvalidation

This procedure will generate a range of residuals from h = 1 to h = 12. Next, we will compute forecast accuracy (MSE) using the colMeans() function in R shown in Table 5.5. The result of this approach is different than the traditional approach whereby Snaive method has the smallest value and thus is the better model. However, Snaive method does not consider trend in forecasting and uses the latest observed value from the past period to forecast the next period. Hence, we will follow the traditional method and use the ETS[M, Ad, M] model for stage 7.

Table 6.2: MSE values of h = 1, 2,, 12

| h = | Snaive | M, A, A | A, A, A | M, Ad, M |
|-----|--------------|--------------|--------------|--------------|
| 1 | 8.102910e+12 | 1.685940e+13 | 1.511931e+13 | 1.003524e+13 |
| 2 | 8.138438e+12 | 3.139975e+13 | 2.828045e+13 | 1.598691e+13 |
| 3 | 8.175078e+12 | 5.382977e+13 | 4.909288e+13 | 2.207597e+13 |
| 4 | 8.221653e+12 | 8.697773e+13 | 8.422110e+13 | 2.768547e+13 |
| 5 | 8.232043e+12 | 1.329465e+14 | 1.295810e+14 | 2.981568e+13 |
| 6 | 8.234395e+12 | 1.792369e+14 | 1.764457e+14 | 3.089319e+13 |
| 7 | 8.291597e+12 | 2.073778e+14 | 2.049681e+14 | 2.405002e+13 |
| 8 | 8.326774e+12 | 2.457916e+14 | 2.442722e+14 | 1.865070e+13 |
| 9 | 8.406946e+12 | 2.811217e+14 | 2.783176e+14 | 1.470019e+13 |

| 10 | 8.480912e+12 | 3.352020e+14 | 3.296131e+14 | 1.070553e+13 |
|----|--------------|--------------|--------------|--------------|
| 11 | 8.558419e+12 | 3.695147e+14 | 3.627407e+14 | 7.457233e+12 |
| 12 | 8.645716e+12 | 3.798052e+14 | 3.709137e+14 | 5.875514e+12 |

Stage 7: Implement Forecasts

We will use the chosen ETS[M, Ad, M] to re-estimate the Pre-Covid parameters.

Table 7.1: Parameters Estimates for ETS Models for Pre-Covid Dataset

| ETS[M, Ad, M] | | | | | | |
|----------------------|----------------|---|---|--|--|--|
| Smoothing parameters | | Initial states | | | | |
| α (level) | 0.352 | lt (level) | 65364573.4781 | | | |
| β (trend) | 0.0264 | bt (slope) | 143164.8325 | | | |
| γ (seasonal) | 1e-04 = 0.0001 | s0, s-1, s-2, s-3, | 0.9733, 0.9149, 0.9893, 0.9472, | | | |
| ф (damped) | 0.9777 | s-4, s-5, s-6, s-7, s-8, s-9, s-10, s-11 | 1.1211, 1.1668, 1.1089, 1.0447, 0.9853, 1.0293, 0.8209, 0.8982 | | | |

In comparison to the parameters in Table 5.2, there are only slight change to the parameter values. In this case, 35% of the forecast is based on the most recent observations, and 65% is based on the previous observations. Only 2% of the slope is based on the most current slope, so a high majority of 98% is based on the most previous slope. The seasonal component value changes very little over time since its value is very small (0.0001). Additionally, the damping parameter (ϕ) is 0.9777 which is almost 1, meaning that the damping is low.

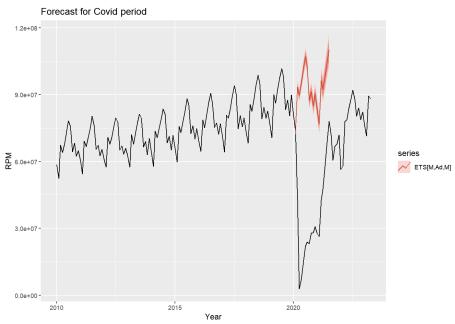


Figure 7.1: Forecast for the Covid period (February 2020 to July 2021 – 18 observations)

As shown in Figure 7.1, it is expected that the forecast will follow the seasonality pattern and trend based on past values. Additionally, the prediction interval is narrow meaning that the variation of possible values is small. Hence, we can say that there is high confidence that the forecasted values will be close to the actual value. Although it is an accurate prediction, the actual data deviates significantly from the forecast. This is due to the unforeseen COVID-19 pandemic which has had an adverse impact on the revenue passenger miles.

Stage 8: Quantifying the Forecasted Lost in Revenue Passenger Miles

Due to the fact that COVID-19 resulted in a drastic fall in the RPM, there is a forecasted value in the revenue passenger value. The forecasted loss is displayed as the huge gap between the forecasted RPM values and the actual RPM values during COVID-19. Using the results we have obtained, a total of approximately 1.035 billion miles of forecasted loss. By using the upper and lower 95% prediction interval, we can calculate the range of the possible forecasted loss. The range of forecasted loss is approximately between 0.957 billion to 1.113 billion miles.

In conclusion, the COVID-19 pandemic has indeed caused a detrimental impact not only on people's health but also on the aviation industry. We have generated a few methods/models to be tested and trained based on the past RPM values before Covid, which then one model is picked based on the fitted values and accuracy measures. This model is then used to forecast the RPM value for the COVID-19 period, in which we expect the trend to go upwards with a strong seasonality pattern. But in reality, the unexpected COVID-19 caused the RPM in the US to plummet drastically to being the lowest ever seen since 2000, resulting in a huge forecasted loss of a total of 1.035 billion miles.

Facing this unpredictable situation is extremely challenging and to better prepare for a similar devastating event in the future, the government should conduct and develop a comprehensive risk management policy outlining strategies for identifying, evaluating, and minimising risks. Additionally, contingency funds or reserves should be established to cover unforeseen future losses. Cost-control measures should also be implemented without compromising the integrity or standard of fundamental business operations.

References:

- Amadeo, K. (2021, December 31). The 2001 recession. *The Balance*. https://www.thebalancemoney.com/2001-recession-causes-lengths-stats-4147962
- Blalock, G., Kadiyali, V., & Simon, D. (2007, November). The Impact of Post-9/11 Airport Security Measures on the Demand for Air Travel. The Journal of Law and Economics, 50(4), 731–755. https://doi.org/10.1086/519816
- Cederholm, T. (2014, November 1). Why are revenue passenger miles important? Yahoo Finance. Retrieved September 12, 2023, from https://finance.yahoo.com/news/why-revenue-passenger-miles-important-170025132.html
- Kenton, W. (2022). Revenue Passenger Mile (RPM): defining a transportation metric. *Investopedia*.

 https://www.investopedia.com/terms/r/revenue-passenger-mile-rpm.asp#:~:text=Kirsten%20Rohrs%20Schmitt-,What%20Is%20Revenue%20Passenger%20Mile%3F,passengers%20by%20the%20distance%20traveled.
- National Academy of Sciences. (2021, August 11). *How has air service changed over time? Air service development and regional economic activity*. Air Service Development and Regional Economic Activity. Retrieved September 12, 2023, from https://crp.trb.org/acrpwebresource12/understanding-air-service-and-regional-economic-activity/how-has-air-service-changed-over-time/
- O'Hare, & Hardingham-Gill. (2020, March 20). Coronavirus: Which countries have travel bans? *CNN Travel*. Retrieved September 12, 2023, from https://edition.cnn.com/travel/article/coronavirus-travel-bans/index.html