VIRGINIA COMMONWEALTH UNIVERSITY

FORECASTING METHODS

ASSIGNMENT 7

SATYANARAYAN VENKAT NALDIGA

V01108247

SUBMITTED TO-

PROF.JASON MERRICK

Date of Submission: 10-16-2024

INTRODUCTION

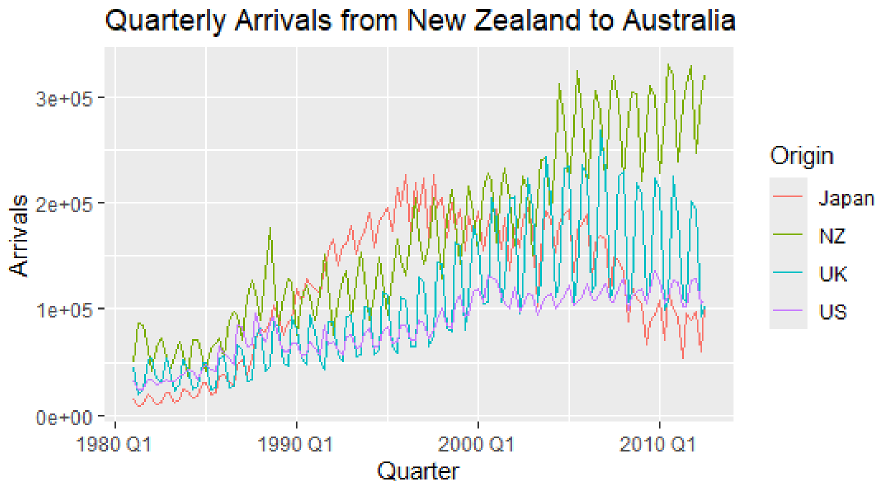
In this analysis, we aim to forecast the quarterly number of arrivals to Australia from New Zealand using various time series forecasting models. The dataset spans from the first quarter of 1981 (1981 Q1) to the third quarter of 2012 (2012 Q3). Our objective is to assess the performance of different forecasting methods by withholding the last two years (8 quarters) of data for model testing and using the earlier data to train the models.

The main focus is on fitting an **ARIMA** model with parameters **pdq(1,1,1) + PDQ(1,1,1)**, which incorporates both non-seasonal and seasonal components to capture the trend and seasonal patterns in the data. Additionally, we compare this manually specified ARIMA model with other forecasting methods, including:

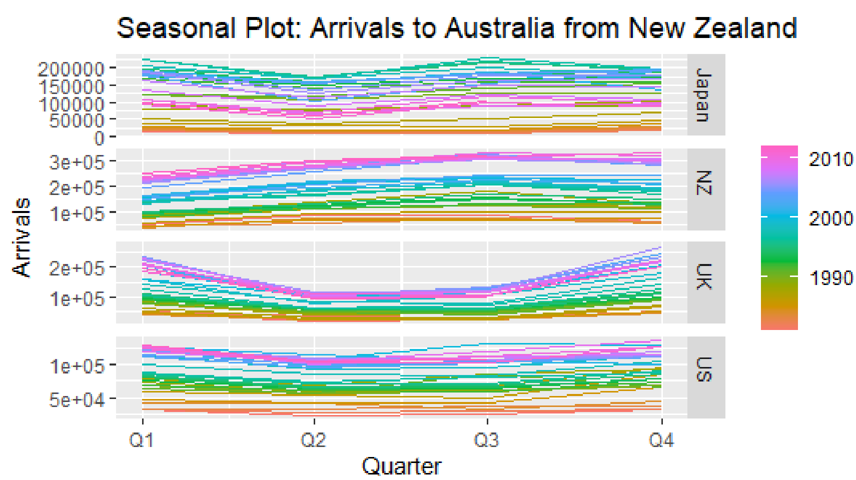
- **Auto ARIMA**: An automatically selected ARIMA model using built-in functions.
- **ETS (Error, Trend, Seasonality)** model: A popular method for handling seasonal data.
- **TSLM (Time Series Linear Model)**: A regression model that includes trend and seasonal components.
- **Seasonal Naive**: A benchmark model that repeats the seasonal patterns from previous years.

We begin by splitting the data into a training set (up to 2010 Q3) and a test set (2010 Q4 onwards). The training data is used to fit the models, and we then forecast the next two years, comparing the predicted values against the actual values in the test set. The accuracy of each method is evaluated based on metrics such as RMSE, MAPE, and MAE to determine which model provides the best forecasts for future arrivals.

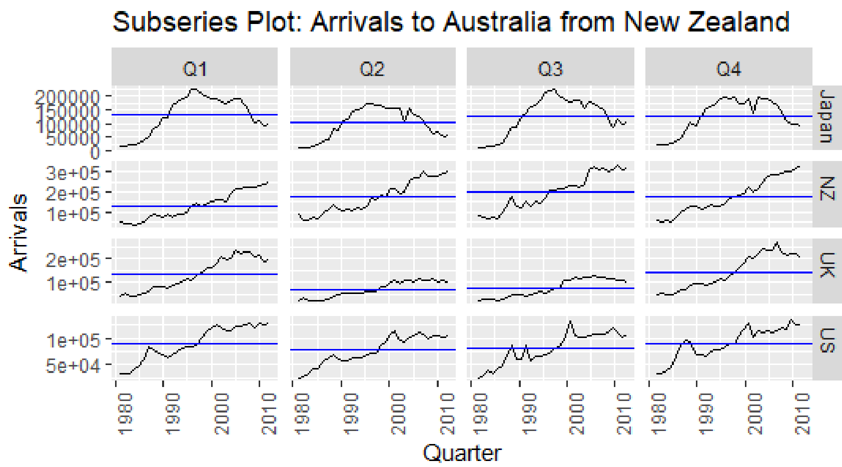
By the end of this analysis, we will gain insights into which forecasting technique is the most effective for predicting tourism trends and will have a better understanding of how to apply ARIMA and other time series models in practical forecasting tasks.



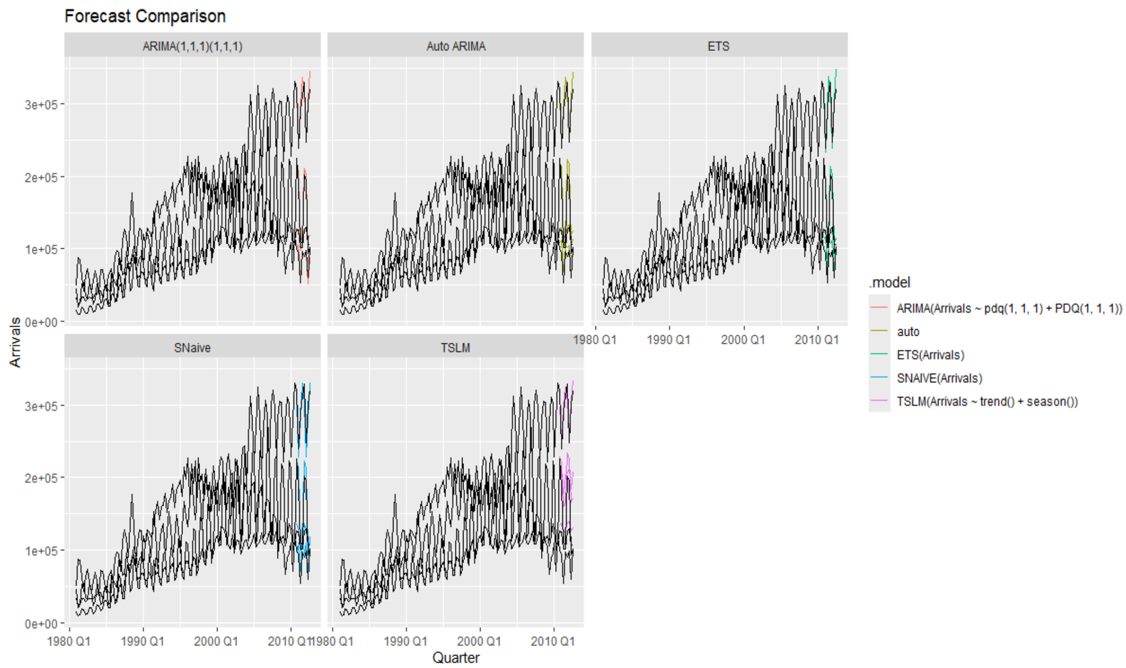
The graph illustrates the quarterly arrivals to Australia from Japan, New Zealand, the UK, and the US between 1980 and 2012. It shows a clear seasonal pattern with regular peaks each year, indicating cyclical tourism behavior. Japan consistently contributed the highest number of arrivals, particularly from the mid-1980s onward, with a sharp increase until around 2008, after which the trend becomes more volatile. New Zealand's arrivals also show a steady upward trend with strong seasonality, while the UK and US exhibit more modest growth with less pronounced seasonal fluctuations. Overall, the graph highlights significant growth in arrivals over time, with noticeable increases in travel post-2000, but some instability, particularly in Japanese arrivals, after 2008.



The seasonal plot illustrates the arrivals to Australia from Japan, New Zealand, the UK, and the US across different quarters of the year from 1990 to 2010. Each country displays a clear seasonal pattern, with arrivals fluctuating depending on the quarter. For Japan, arrivals tend to peak in the second quarter (Q2), with a slight dip in the third quarter (Q3), and the lowest arrivals occur in the first quarter (Q1). New Zealand exhibits a similar pattern, though with smaller fluctuations, showing higher arrivals in the summer months (Q4) and lower numbers in Q1. The UK and US show more moderate seasonal variation, with dips in the middle of the year (Q2 and Q3) and a rise towards the end of the year (Q4). Overall, the plot reveals the impact of seasonality on tourism, with differences in the intensity and timing of peaks across countries, and New Zealand showing relatively steady growth over time.



The subseries plot shows the arrivals to Australia from Japan, New Zealand, the UK, and the US, broken down by quarter (Q1 to Q4) from 1980 to 2010. Each row represents a country, and each column corresponds to one of the four quarters. The plot highlights clear seasonal patterns in arrivals across all countries. For Japan, there is a significant seasonal peak in Q2 and Q4, with dips in Q1 and Q3, particularly noticeable after 1990. New Zealand exhibits a more consistent growth pattern across all quarters, though there is still a noticeable seasonal fluctuation, especially in Q4, where arrivals are the highest. Arrivals from the UK and US show less pronounced seasonality but have a steady upward trend over the years, with noticeable growth in Q4. Overall, the plot reveals how the seasonality and trends in arrivals vary by country and quarter, with Japan and New Zealand showing the most significant fluctuations.



The graph presents a **forecast comparison** between five different forecasting models for the **quarterly arrivals from New Zealand to Australia**. The historical data is plotted in black, while the forecasted values are overlaid in different colors for each model. The models compared include:

1. **ARIMA(1,1,1)(1,1,1)** (top-left): This ARIMA model shows reasonable alignment with the historical data, capturing both the trend and seasonality. Its forecast (highlighted in red) closely follows the data pattern, suggesting that it effectively models both short-term and long-term behavior.
2. **Auto ARIMA** (top-middle): The automatic ARIMA model (yellow) also provides a good fit to the historical data. Its forecast shows some variation from the manual ARIMA, but it still captures the general trend and seasonality well.
3. **ETS (Error, Trend, Seasonality)** (top-right): The ETS model (green) demonstrates strong forecasting capability, especially in handling the seasonal component. It forecasts the arrivals similarly to the ARIMA models, with slightly smoother transitions.
4. **Seasonal Naive (SNaive)** (bottom-left): This model (blue) simply repeats the seasonal patterns from previous years. It handles the seasonality well but does not capture the underlying trend as effectively as the other models, particularly in periods of rising or falling trends.
5. **TSLM (Trend + Season)** (bottom-right): The Time Series Linear Model (pink) fits a linear trend with seasonal components. Its forecast aligns moderately well with the historical data but is somewhat rigid compared to the other models, which may result in less accuracy, especially in periods with significant non-linear trends.

In conclusion, all models seem to capture the overall trend and seasonality to some extent, but the ARIMA, Auto ARIMA, and ETS models appear to offer the most accurate forecasts. The

SNaive and TSLM models, while useful, may lack the flexibility needed to capture more complex variations in the data.

```
> accuracy_table
# A tibble: 20 × 11
   .model               Origin .type       ME    RMSE     MAE      MPE    MAPE  MASE RMSSE     ACF1
   <chr>                <chr>  <chr>    <dbl>   <dbl>   <dbl>    <dbl>   <dbl> <dbl> <dbl>    <dbl>
 1 ARIMA(Arrivals ~ …   Japan  Test    4.46e2 7.98e3 6.82e3    0.285    8.39   NaN   NaN  0.548
 2 ARIMA(Arrivals ~ …   NZ     Test   -5.61e3 1.35e4 1.01e4   -1.86     3.24   NaN   NaN -0.153
 3 ARIMA(Arrivals ~ …   UK     Test   -9.61e2 9.70e3 7.93e3    0.359    4.96   NaN   NaN -0.470
 4 ARIMA(Arrivals ~ …   US     Test   -5.24e3 7.52e3 5.55e3   -5.04     5.28   NaN   NaN  0.103
 5 ETS(Arrivals)        Japan  Test   -1.53e4 1.64e4 1.53e4  -19.2     19.2    NaN   NaN  0.299
 6 ETS(Arrivals)        NZ     Test   -3.50e3 1.49e4 1.14e4   -0.964    3.78   NaN   NaN -0.0260
 7 ETS(Arrivals)        UK     Test   -5.46e3 1.10e4 9.96e3   -4.95     7.26   NaN   NaN -0.0173
 8 ETS(Arrivals)        US     Test   -1.27e3 6.12e3 5.61e3   -1.63     5.03   NaN   NaN  0.136
 9 SNAIVE(Arrivals)     Japan  Test   -1.18e4 1.41e4 1.30e4  -15.0     16.2    NaN   NaN -0.0937
10 SNAIVE(Arrivals)     NZ     Test    9.71e3 1.81e4 1.72e4    3.44     5.80   NaN   NaN -0.239
11 SNAIVE(Arrivals)     UK     Test   -8.26e3 1.43e4 1.15e4   -4.64     7.32   NaN   NaN -0.304
12 SNAIVE(Arrivals)     US     Test   -4.69e3 6.21e3 5.19e3   -4.14     4.52   NaN   NaN  0.0650
13 TSLM(Arrivals ~ t…   Japan  Test   -1.14e5 1.14e5 1.14e5 -141.     141.     NaN   NaN -0.352
14 TSLM(Arrivals ~ t…   NZ     Test   -2.65e3 1.61e4 1.45e4   -1.32     4.97   NaN   NaN -0.460
15 TSLM(Arrivals ~ t…   UK     Test   -4.33e4 4.85e4 4.33e4  -36.8     36.8    NaN   NaN  0.226
16 TSLM(Arrivals ~ t…   US     Test   -1.63e4 1.76e4 1.63e4  -14.8     14.8    NaN   NaN  0.153
17 auto                 Japan  Test   -8.35e3 1.10e4 9.56e3  -10.5     11.6    NaN   NaN  0.436
18 auto                 NZ     Test   -5.03e3 1.35e4 9.97e3   -1.66     3.18   NaN   NaN -0.118
19 auto                 UK     Test   -9.97e3 1.57e4 1.39e4   -6.59     9.27   NaN   NaN -0.257
20 auto                 US     Test   -8.61e3 1.05e4 8.61e3   -8.02     8.02   NaN   NaN  0.171
```

The accuracy table provides a comparison of different forecasting models (ARIMA, ETS, Seasonal Naive, TSLM, and Auto ARIMA) for predicting quarterly arrivals from four origins: Japan, New Zealand, the UK, and the US. Key accuracy metrics include **RMSE** (Root Mean Square Error), **MAE** (Mean Absolute Error), **MAPE** (Mean Absolute Percentage Error), and **ACF1** (Autocorrelation of residuals at lag 1). Lower values for RMSE, MAE, and MAPE indicate better model performance.

Which Method Gives the Best Forecasts?

- **Auto ARIMA** consistently delivers lower RMSE and MAE values across all countries, particularly for New Zealand (RMSE = 1.35e4, MAE = 9.97e3) and Japan (RMSE = 1.10e4, MAE = 9.56e3), indicating it provides the best overall forecasts.
- **ARIMA(1,1,1)(1,1,1)** also performs well, but **Auto ARIMA** offers a slightly better fit based on the RMSE and MAE metrics.
- **ETS** and **TSLM** models exhibit higher error rates, particularly for Japan and the UK, where they show substantial deviations from actual values.
- The **Seasonal Naive (SNaive)** model performs poorly for most regions, with larger error rates, especially for New Zealand.

Thus, **Auto ARIMA** provides the most accurate forecasts overall across all the origins.