Assignment 1

Social Forecasting

Due: March 29, 2024

Instructions

- When submitting your answers, please 1) show your code and 2) make sure your presentation is clear and easy to follow. Rmarkdown will make it easy to produce a suitable report, but you are welcome to use another approach if you prefer.

Answer Task 1

```r
# Install packages
install.packages("tidyverse")
install.packages("forecast")
install.packages("zoo")

# Load the libraries
library(tidyverse)
library(forecast)
library(zoo)

#TASK 1
# Import the dataset
rd <- read_csv("/Users/iseli/Downloads/AFG_IRL_refugeeData.csv")
```

Answer Task 2

```r
# Extract the column "refugees under UNHCR mandate" and create the time
    series object
rdts <- ts(rd[,6], start=1999, end=2021, frequency=1)

# Print the time series object
print(rdts)
```

Answer Task 3

```r
# Create the plot using autoplot
plot_rdts <- autoplot(rdts) +
  labs(title = "Time Series of Refugees under UNHCR Mandate",
       x = "Year",
       y = "Number of Refugees") +
  theme_minimal()

# Print the plot --FIRST PLOT--
print(plot_rdts)
```

1 Forecasts vs Actual Data

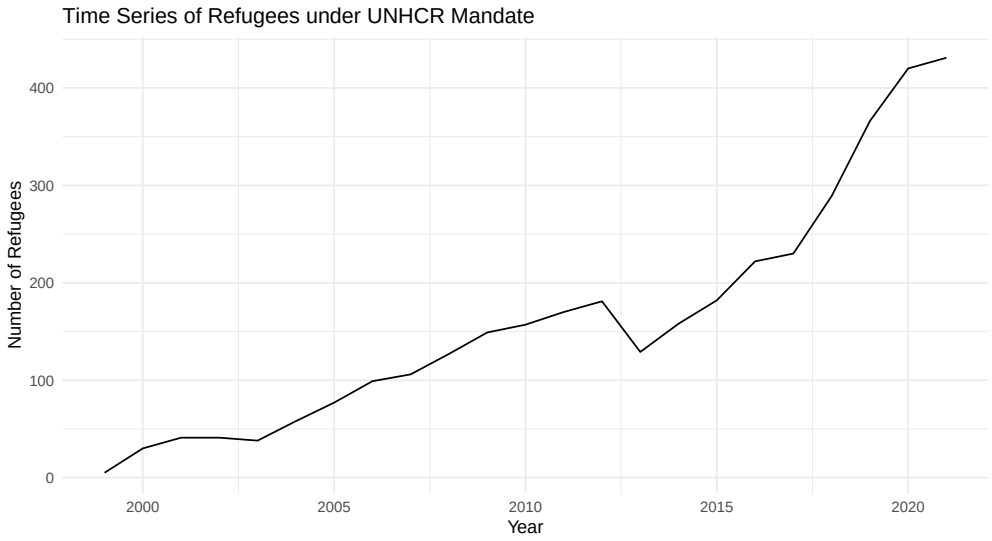

Figure 1: Forecasts vs Actual Data (2016-2021)

Answer Task 4

```r
# Subset the time series to include only data from 1999 to 2015
rdts_subset <- window(rdts, start=1999, end=2015)

# Calculate the mean of the subsetted time series
mean_value <- mean(rdts_subset)

# Create a time series object of the mean value for the forecast years
    2016 to 2021
forecast_values <- rep(mean_value, times=6)
```

```r

# Create a ts object for the forecasted values
forecast_ts <- ts(forecast_values, start=2016, end=2021, frequency=1)

# Print the forecasted values
print(forecast_ts)

# Just making sure forecast package is already loaded
if (!requireNamespace("forecast", quietly = TRUE)) {
  install.packages("forecast")
}
library(forecast)

# Create the naive forecast
naive_forecast <- naive(rdts_subset, h=6)

# Print the forecasted values
print(naive_forecast)
naive_forecast <- forecast::naive(rdts_subset, h=6)$mean

# Just double checking that packages are already loaded
if (!requireNamespace("forecast", quietly = TRUE)) {
  install.packages("forecast")
}
library(forecast)

# rwf() function with drift=TRUE to create the drift forecast
drift_forecast <- rwf(rdts_subset, h=6, drift=TRUE)

# Print the forecasted values
print(drift_forecast)
drift_forecast <- forecast::rwf(rdts_subset, h=6, drift=TRUE)$mean

# Extract the actual data for the years 2016 to 2021
actual_data <- window(rdts, start=2016, end=2021)

#Encountering many errors, -I will redo some codes:
# Subset the time series to include only data from 1999 to 2015
rdts_subset <- window(rdts, start=1999, end=2015)

# Calculate the average forecast
average_value <- mean(rdts_subset)
average_forecast <- rep(average_value, times=6)
average_forecast_ts <- ts(average_forecast, start=2016, end=2021,
    frequency=1)

# Calculate the naive forecast
naive_forecast <- forecast::naive(rdts_subset, h=6)
naive_forecast_values <- naive_forecast$mean

# Calculate the drift forecast
```

```r
drift_forecast <- forecast::rwf(rdts_subset, h=6, drift=TRUE)
drift_forecast_values <- drift_forecast$mean
```

Answer Task 5

```r
# Extract the actual data for the years 2016 to 2021 from the original time series object
actual_data <- window(rdts, start=2016)

# Combine all forecasts and actual data into a data frame for plotting
combined_data <- data.frame(
  Year = rep(2016:2021, times=4),
  Value = c(average_forecast_ts, naive_forecast_values, drift_forecast_values, actual_data),
  Method = rep(c("Average", "Naive", "Drift", "Actual"), each=6)
)

# Load ggplot2 for plotting
library(ggplot2)

# Create the combined plot
combined_plot <- ggplot(combined_data, aes(x = Year, y = Value, color = Method, group = Method)) +
  geom_line() +
  geom_point() +
  labs(title = "Forecasts vs Actual Data (2016-2021)",
       x = "Year",
       y = "Number of Refugees",
       color = "Method") +
  theme_minimal()

# Print the combined plot -SECOND PLOT-
print(combined_plot)

#Interpretation of plot
#The actual data line shows the real number of refugees from Afghanistan to Ireland for each year from 2016 to 2021
#By comparing the actual data with the forecasts, it's evident that there are variations in the actual number of refugees over these years
#The actual data line allows us to evaluate the accuracy of each forecasting method.
#Based on the plot, it appears that none of the forecasting methods perfectly match the actual data, which is expected in real-world scenarios due to the complexity and unpredictability of factors influencing refugee movements
```

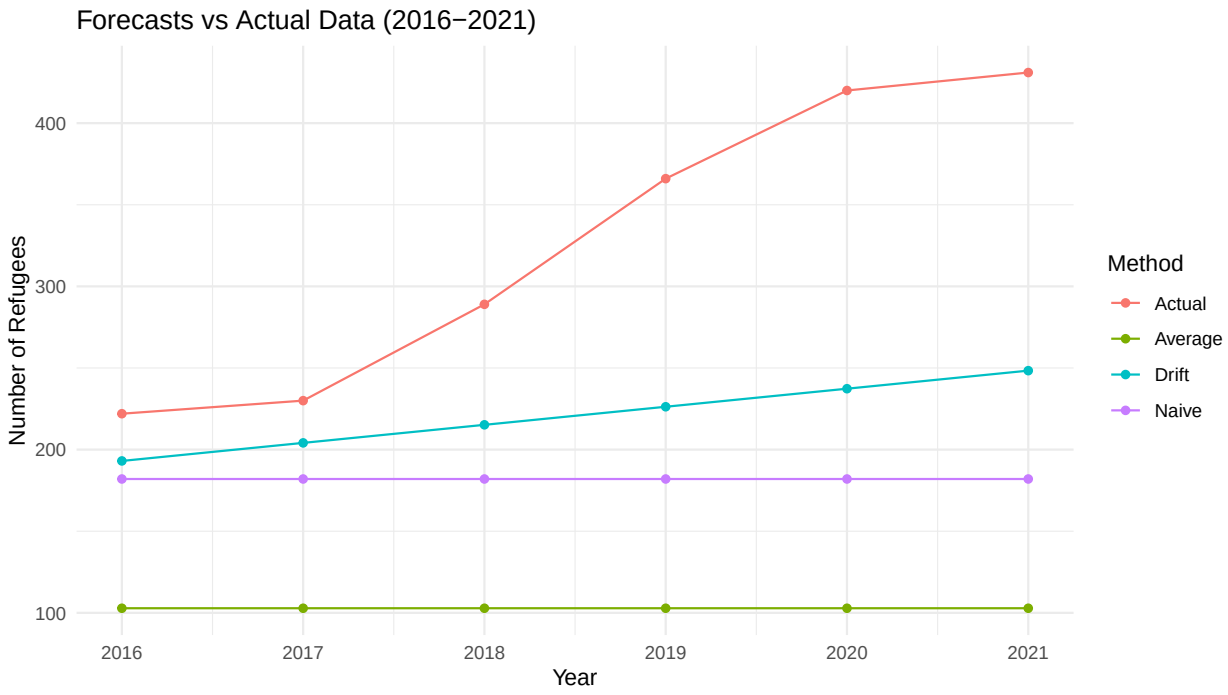

Figure 2: Forecasts vs Actual Data (2016-2021)

2 Forecasts vs Actual Data

As shown in Figure 2, the drift method provides the closest approximation to the actual data from 2016 to 2021.

Answer Task 6

```
#The Drift method, might offer a slightly more flexible approach compared
    to the static forecasts of the Average and Naive methods.
#I would not particullarily suggest an improvement. The effectiveness of
    each method would depend on how well it aligns with the actual
    observed trends in the data.
```