Assignment 2(TS)

Gavin Kunish

2025-06-30

Set Up

```
setwd('/Users/gavin-kunish/Desktop')
library(tsibble)
```

```
## Registered S3 method overwritten by 'tsibble':
##   method               from
##   as_tibble.grouped_df dplyr
```

```
##
## Attaching package: 'tsibble'
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, union
```

```
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

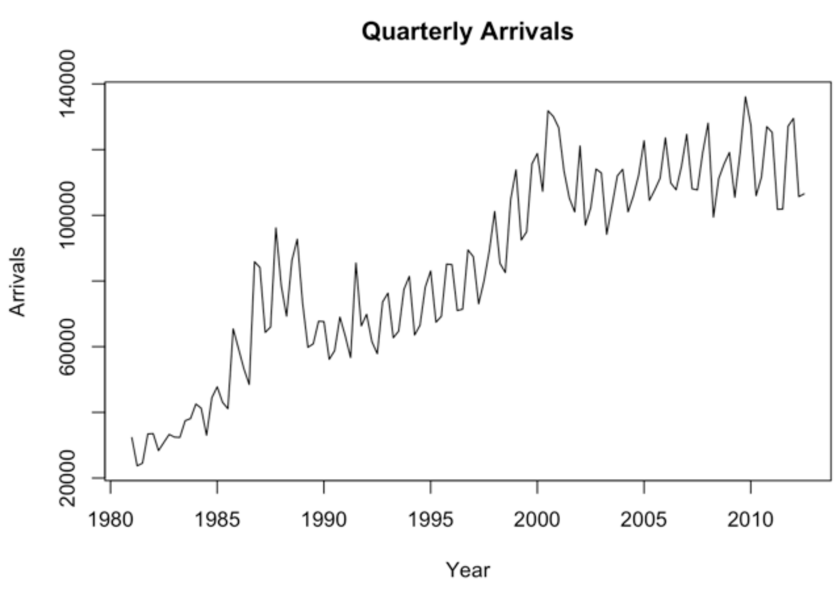
```
library(forecast)
load("visitors.rda")

#str(visitors)
head(visitors)
```

```
## # A tsibble: 6 x 3 [1Q]
## # Key:       Origin [1]
##   Quarter Origin Arrivals
##     <qtr> <chr>     <int>
## 1 1981 Q1 US        32316
## 2 1981 Q2 US        23721
## 3 1981 Q3 US        24533
## 4 1981 Q4 US        33438
## 5 1982 Q1 US        33527
## 6 1982 Q2 US        28366
```

```
visitors_ts <- ts(visitors$Arrivals, frequency = 4, start = c(1981, 1))
```

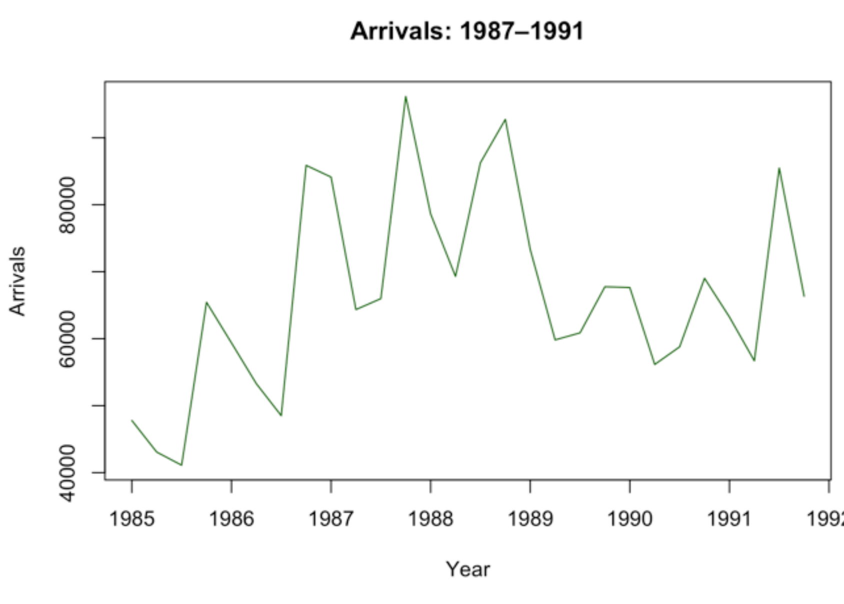
Question 1

```
plot(visitors_ts, main = "Quarterly Arrivals", ylab = "Arrivals", xlab = "Year")
```



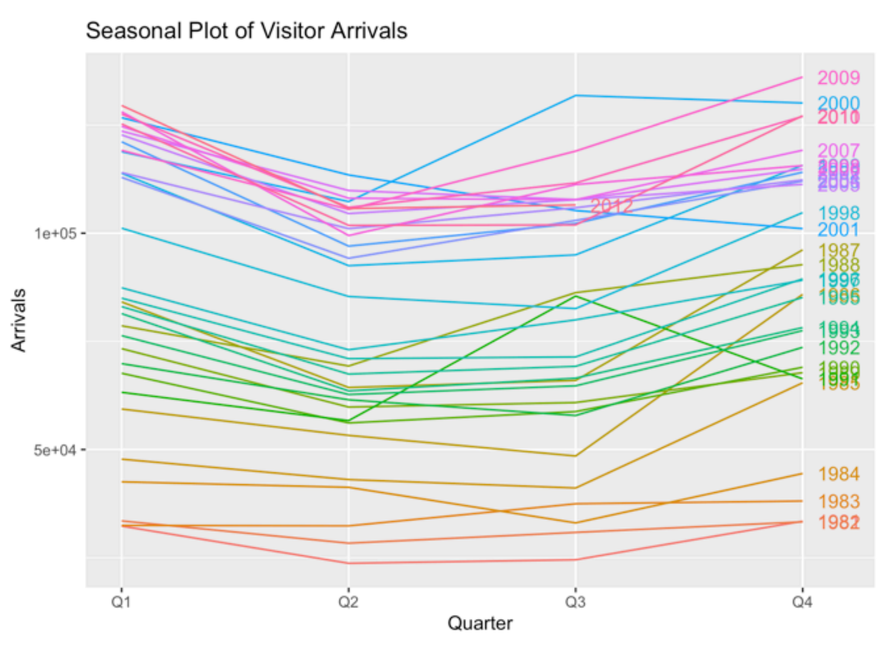
There is a definitive upward trend throughout the entire time series. We also see strong seasonality with higher values in Q2 and Q3. There also appears to be increasing variability during the time series suggesting a multiplicative model. Notably there are large spikes in the mid/late 1980's and early 2000s, with some flattening in the late 2000's.

```
plot(window(visitors_ts, start = c(1985, 1), end = c(1991, 4)),
     main = "Arrivals: 1987–1991", ylab = "Arrivals", xlab = "Year", col = "darkgreen")
```

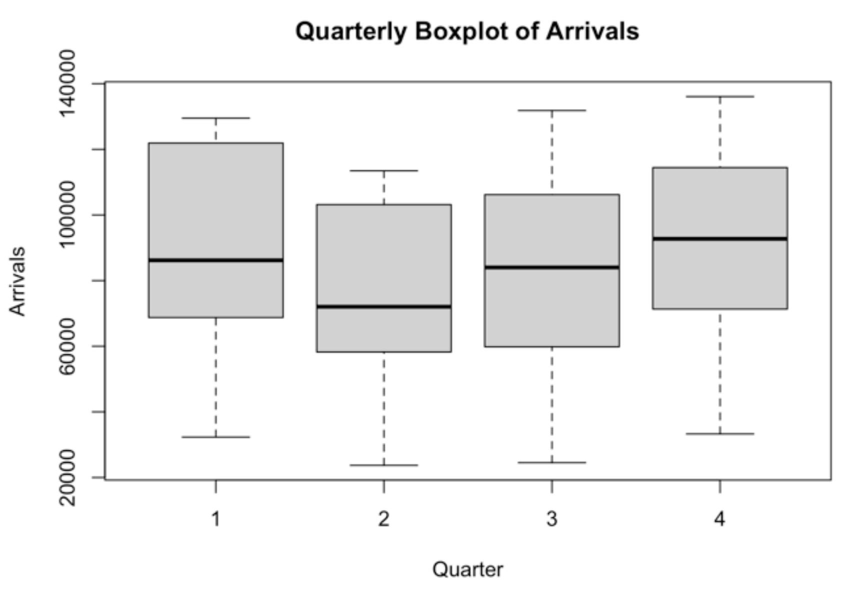


After some research, the US's economy was strong in the mid 1980's as well as the passing of the Airline Deregulation Act, which made fares cheaper and offered more airline routes. These could explain the sudden increase in airline arrivals. Conversly we see a drop off in ~1989. This may be attributed to "Black Monday" which occured in late 1987 and potentially had a lagging effect on overseas travelers. The spike in arrivals in the early 2000's may have something to do with the dot com boom in the late 1990's. The sudden drop in arrivals and flattening of growth may be attributed to the 9/11 attacks. Other noteworthy events could be the olympics, which were held in the United States in 1984 and 1996, both directly before these large increases in arrivals.

```
ggseasonplot(visitors_ts, year.labels = TRUE,
             main = "Seasonal Plot of Visitor Arrivals", ylab = "Arrivals")
```




```
boxplot(visitors_ts ~ cycle(visitors_ts), main = "Quarterly Boxplot of Arrivals", xlab = "Quarter", ylab = "Arriv
als")
```



From these two seasonal plots, we can see that quarter 2 has the lowest arrivals across most the years. Quarters 1 and 3 appear to be similar over the years with a slight trend towards Quarter 1 being higher in the more recent years. Quarter 4 also is consistantly the highest for arrivals. We can confirm these through the seasonal boxplots. Quarter 4 has the highest median and Quarter 3 has the widest range while having a similar median to Quarter 1. We can also see that over time the arrivals are increasing. We can assume from these plots that the time series does have seasonality and that these seasonal effects increase over time, leading us to believe it is a multiplicative model.

Question 2

```
adf.test(visitors_ts)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  visitors_ts
## Dickey-Fuller = -2.4609, Lag order = 5, p-value = 0.385
## alternative hypothesis: stationary
```

As we assumed, our p-value is substantially greater than .05, meaning this time series is non-stationary, and we can use the Holt-Winters method. Given all the information we have, a multiplicitive Holt-Winters model would be the most optimal to use. It is non-stationary and seasonality increases with level.

Question 3

```
additive_fit <- hw(visitors_ts, seasonal = "additive", h = 14)

mult_fit <- hw(visitors_ts, seasonal = "mult", h = 14)

add_damp_fit <- hw(visitors_ts, seasonal = "additive", damped = T, h = 14)

mult_damp_fit <- hw(visitors_ts, seasonal = "mult", damped = T, h = 14)

mult_exp_fit <- hw(visitors_ts, seasonal = "mult", exponential = T, h = 14)
```

Question 4

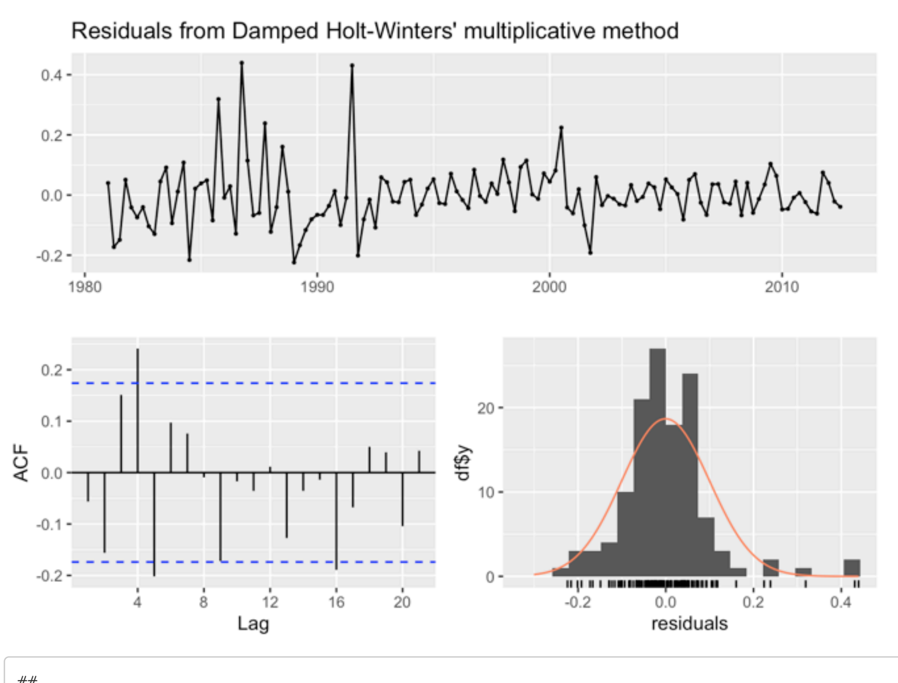
```
add_test <- accuracy(additive_fit)

mult_test <- accuracy(mult_fit)

add_damp_test <- accuracy(add_damp_fit)

mult_damp_test <- accuracy(mult_damp_fit)

mult_exp_test <- accuracy(mult_exp_fit)

model_names <- c("Additive",
                 "Multiplicative",
                 "Additive + Damped",
                 "Multiplicative + Damped",
                 "Exp Trend + Mult Season")

rmse_values <- c(add_test[2],
                 mult_test[2],
                 add_damp_test[2],
                 mult_damp_test[2],
                 mult_exp_test[2])

results <- data.frame(Model = model_names, RMSE = rmse_values)
print(results)
```

```
##                     Model     RMSE
## 1                Additive 7542.656
## 2          Multiplicative 7550.956
## 3       Additive + Damped 7552.064
## 4 Multiplicative + Damped 7460.002
## 5 Exp Trend + Mult Season 7870.123
```

Based on these results, the Multiplicative and Damped Holt-Winter Model is the best as it has the lowest RMSE. This makes sense because we have the smoothing of the time series near the end, and the damping helps to temper our forecasted trend.

Question 5

```
checkresiduals(mult_damp_fit)
```




```
##
##  Ljung-Box test
##
## data:  Residuals from Damped Holt-Winters' multiplicative method
## Q* = 21.921, df = 8, p-value = 0.005064
##
## Model df: 0.   Total lags used: 8
```

Our residual plot does mainly fluctuate around 0 with some spikes prior to ~1993. However it does smooth out post 1993. This suggests the model is a good fit overall, and an even better fit on more recent data. With our Ljung-Box returning a p-value well below .05, we can confirm that the residuals are not white noise and there is still autocorrelation that the model is unable to capture.

Next with our ACF Plot, we can further confirm that the residuals are not white noise because the lags are not centered around 0 and that there is still some patterns not captured by the model because of the spikes in the lags exceeding our significance threshold.

Finally the histogram of the residuals leave something to be desired. The residuals do not appear to be normal. There is a clear right skew and the peaks of the resiudals don't exactly conform to normality as they are very high and slightly skewed.

```
summary(mult_damp_fit)
```

```
##
## Forecast method: Damped Holt-Winters' multiplicative method
##
## Model Information:
## Damped Holt-Winters' multiplicative method
##
## Call:
##  hw(y = visitors_ts, h = 14, seasonal = "mult", damped = T)
##
##   Smoothing parameters:
##     alpha = 0.52
##     beta  = 0.0027
##     gamma = 1e-04
##     phi   = 0.98
##
##   Initial states:
##     l = 26914.2591
##     b = 2002.8599
##     s = 1.0638 0.9467 0.9133 1.0762
##
##   sigma:  0.1034
##
##      AIC     AICc      BIC
## 2913.642 2915.538 2942.084
##
## Error measures:
##                     ME     RMSE     MAE        MPE     MAPE      MASE
## Training set -1.834419 7460.002 5363.22 -0.9701216 6.852661 0.7266756
##                      ACF1
## Training set -0.001971966
##
## Forecasts:
##         Point Forecast     Lo 80    Hi 80    Lo 95    Hi 95
## 2012 Q4       122225.2 106033.32 138417.1 97461.86 146988.5
## 2013 Q1       123798.8 105276.90 142320.8 95471.97 152125.7
## 2013 Q2       105187.4  87816.96 122557.8 78621.62 131753.1
## 2013 Q3       109171.5  89578.75 128764.3 79206.97 139136.1
## 2013 Q4       122829.1  99138.55 146519.6 86597.54 159060.7
## 2014 Q1       124397.5  98830.37 149964.7 85295.93 163499.2
## 2014 Q2       105685.3  82691.88 128678.6 70519.92 140850.6
## 2014 Q3       109677.3  84552.31 134802.3 71251.94 148102.7
## 2014 Q4       123386.1  93754.23 153018.0 78068.05 168704.2
## 2015 Q1       124949.8  93607.44 156292.1 77015.81 172883.7
## 2015 Q2       106144.5  78421.35 133867.7 63745.60 148543.4
## 2015 Q3       110143.9  80270.02 140017.7 64455.77 155832.0
## 2015 Q4       123899.9  89083.75 158716.1 70653.18 177146.7
## 2016 Q1       125459.1  89009.06 161909.2 69713.57 181204.7
```

Our Alpha value of 0.52 puts moderate weight on recent levels which means it can adapt fairly well to level shifts.

Our Beta value of .0027 leads to a low level of trend learning which means it wont adapt to sudden new trends very quickly.

Our Gamma value of .0001 fixes seasonality, so pattern shifting over the quarters wont be captured.

Our Phi value of .98 slows the damping.

Question 6

```
naive_fit <- snaive(visitors_ts, h = 14)

naive_test <- accuracy(naive_fit)

rmse_mult <- mult_damp_test[2]
rmse_naive <- naive_test[2]

new_results <- data.frame(
  Model = c("Holt-Winters (Multiplicative + Damped)", "Seasonal Naive"),
  RMSE = c(rmse_mult, rmse_naive)
)
print(new_results)
```

```
##                                    Model      RMSE
## 1 Holt-Winters (Multiplicative + Damped)  7460.002
## 2                         Seasonal Naive 10298.985
```

The Holt-Winters model with damping significantly outperforms the Seasonal Naive model when comparing their RMSE's.

Question 7

a.

The first plot in the decomposition shows the total number of people in the Australian labor force, otherwise known as the value plot. In this plot we can see a strong upward trend from ~6000 at the beginning, and ending around 9000.

In the second plot, or the trend plot, we see a steady and smooth increase over time. Since the trend plot ignores small fluctuations and seasonality, and the value plot doesn't have any large spikes, it's acting exactly as we would expect. The trend plot also follows the same scaling as the value plot.

The third plot is the seasonality plot. In this we see a very small increase in variability over time as the range goes from (~100, -75) to (over 100, -100). We also notice that the scale of the seasonal plot changes from the first two. This leads us to belive that while the seasonality is present, it is fairly modest.

The final plot is of the remainders/residuals. This plot shows us fluctuation that is not addressed by seasonality or trend. This plot also mostly operates inbetween 100 and -100 like the seasonality plot. However we do see one large deviation in the early 1990's where we have a large negative spike all the way down to -400.

b.

Yes, the 1991/1992 Austrailian recession is visible in these plots. This would line up with the large negative spike in the residual plot explained in part a. We can also see a dip in the value plot at the same time, as well as a bit of flattening in the trend line.