

Sanchez Jarrett 20109664

Jarrett Sanchez

2023-10-08

Question 1 - ETS (20 marks)

0) Scale the data (e.g., divide by 100). From now on, you'll work with the scaled series (0 marks).

```
# read csv file
aus_steroids <- read.csv("H03_drug_Sales_Australia.csv")

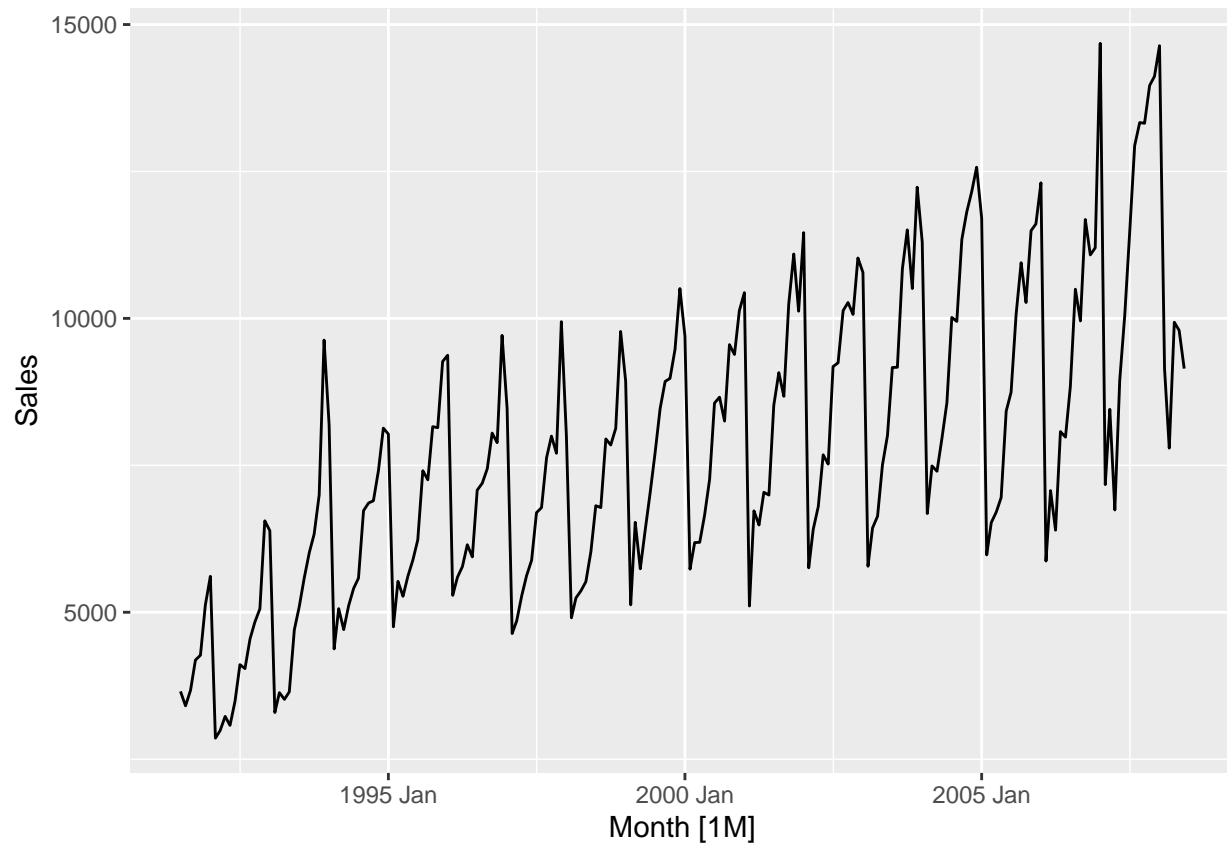
# scale data, convert to time series
steroids <- aus_steroids |>
  mutate(Month = yearmonth(Period), Sales = Sales/100) |>
  select(-Time, -Period) |>
  as_tsibble(index = Month)

steroids
```

```
## # A tsibble: 204 x 2 [1M]
##   Sales      Month
##   <dbl>    <mth>
## 1 3653. 1991 Jul
## 2 3408. 1991 Aug
## 3 3673. 1991 Sep
## 4 4187. 1991 Oct
## 5 4270. 1991 Nov
## 6 5123. 1991 Dec
## 7 5611. 1992 Jan
## 8 2858. 1992 Feb
## 9 2986. 1992 Mar
## 10 3228. 1992 Apr
## # i 194 more rows
```

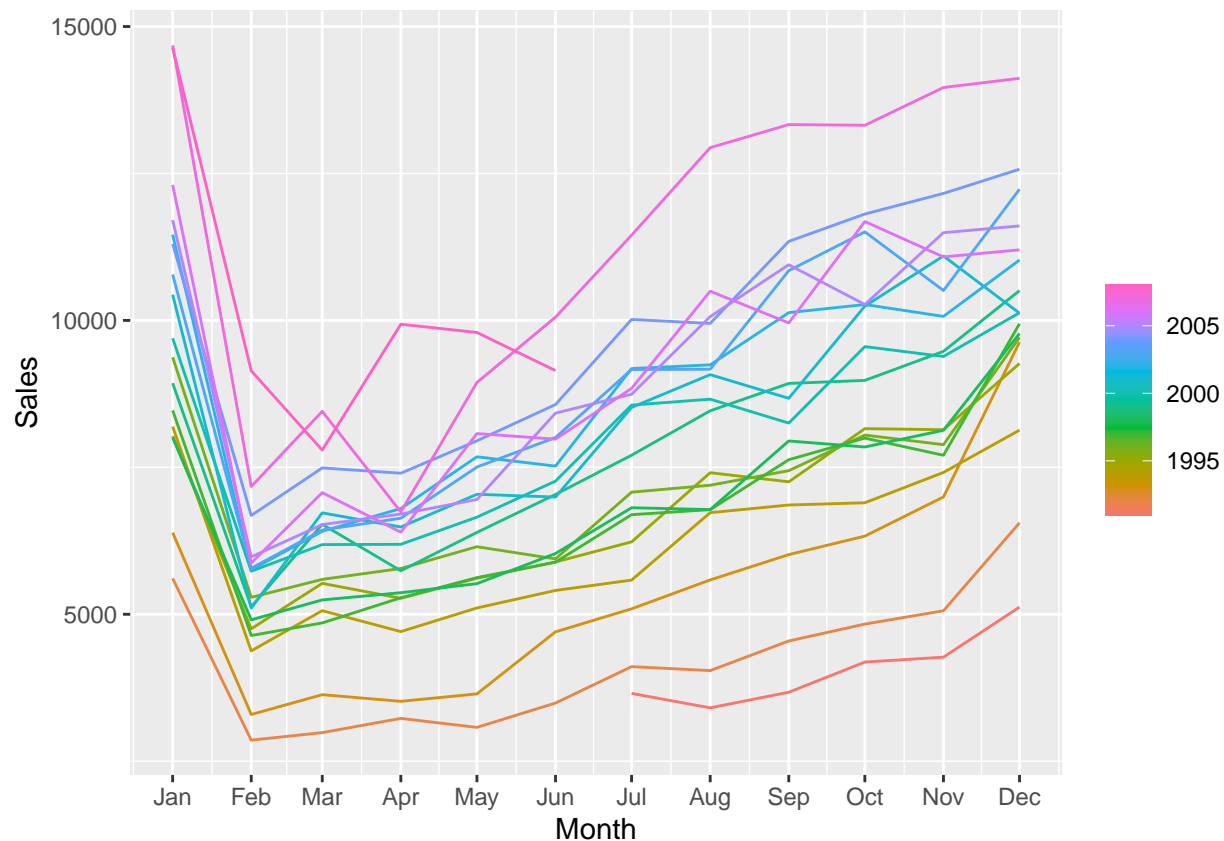
a) Plot the series and discuss the main features, including stationarity (2 marks).

```
steroids |>
  autoplot(Sales)
```



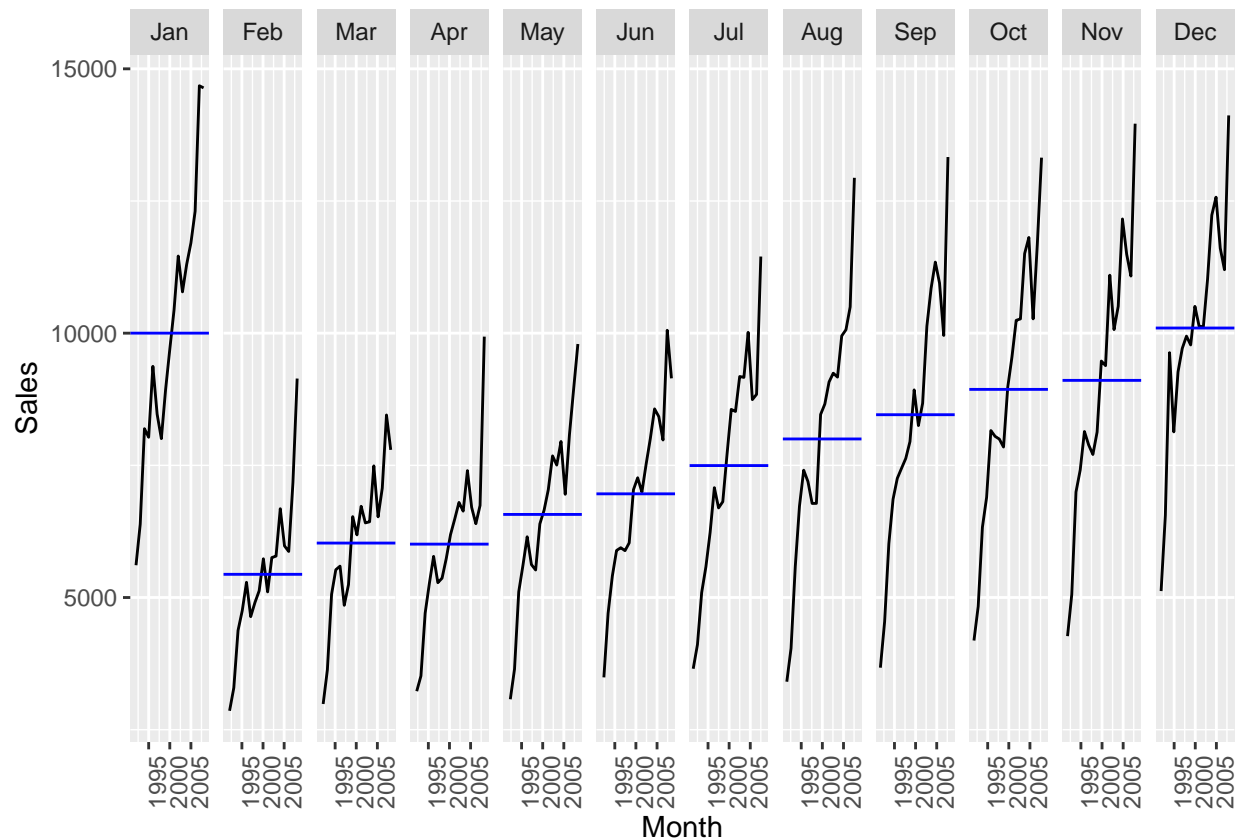
The series looks non-stationary. There is a clear upwards trend and a strong seasonal pattern. Variability in the data is non-constant, where the variance appears proportional to the level of series over time.

```
steroids |>  
  gg_season(Sales)
```



The plot above confirms seasonality. There appears to be a drop in H03 sales around February of each year. This is likely due to Australia's Pharmaceutical Benefits Scheme which subsidises medicine, making it easier for patients to stockpile them at the end of the year. The Scheme's co-payment amount changes on 1 January each year, which explains the increase in sales towards December as patients would like to stockpile medicine before any potential price increase occurs.

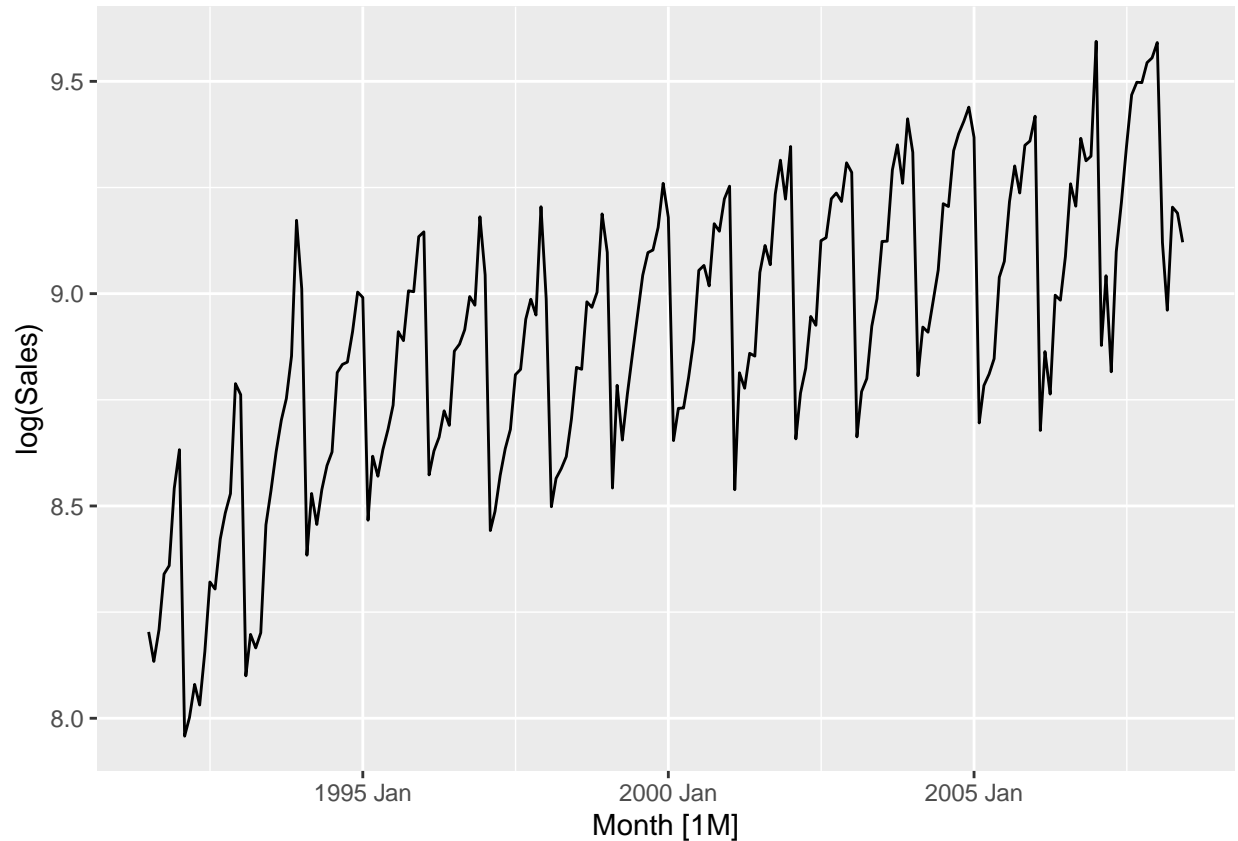
```
steroids |>
  gg_subseries(Sales)
```



Strong trend in all months, largest trend in January, larger increase in second half of the year (July to December) compared to first half (excluding January).

b) Forecast next two years using SES, Holt's linear and Holt's damped trend. Plot the series and the forecasts. Merely based on this plot, discuss the adequacy of these methodologies to forecast from this series. Explain your answer (5 marks total).

```
# stabilise variance
steroids |>
  autoplot(log(Sales))
```



- c) Repeat b) with Holt-Winters' seasonal methods. Discuss whether additive or multiplicative seasonality is necessary. Explain your answer (5 marks total).
- d) Compare MSE and MAE of one-step-ahead, four-step-ahead, and six-step-ahead forecasts from methods discussed in b and c above. Report your results neatly and clearly. You can use a Table. Which method has the highest accuracy? Does this selection depend on the number of pre-specified (steps-ahead) forecasts? Explain your answer (5 marks total).
- e) Briefly discuss the potential mistake/error we may unintentionally introduce in the discussion when comparing models b and c using the MSE and MAE. (3 marks)